# Improving Predictive Accuracy in Early Disease Detection Using Hybrid Neural Network Architectures and Feature Engineering

Sanjay Kumar Pandey[1†], Dharmendra Kumar[2*], Bechoo Lal[3†]

[1]Computer Science , Shri Jagdishprasad Jhabarmal Tibrewala University, Vidyanagri, Jhunjhunu Bisau Road, Chudela,, Jhunjhunu, 333010, Rajasthan, India.
[2*]Computer Science and Engineeing, United College of Engineering and Research, Naini, Prayagraj, 211010, Uttar Pradesh, India.
[3]Computer Science , Shri Jagdishprasad Jhabarmal Tibrewala University, Vidyanagri, Jhunjhunu Bisau Road, Chudela,, Jhunjhunu, 333010, Rajasthan, India.

†These authors contributed equally to this work.

**Abstract**

This research focuses on enhancing early disease detection through a novel hybrid neural network (HNN) framework, integrating Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks. The HNN effectively handles diverse medical data, encompassing spatial information from medical imaging and temporal patterns from clinical time-series records, to improve diagnostic accuracy. The general methods of diagnosis have had problems related to low sensitivity and specificity, particularly when used to diagnose economic diseases at their initial stage. By combining the high performance associated with CNNs in image pattern recognition and the LSTMs in sequence modeling, the presented HNN model contributes to the improvement in the accuracy and interpretability of early identification of diseases such as cancer and cardiovascular diseases. Feature engineering is used to obtain specific characteristics from certain sources, including the texture and morphology of the medical image and the heart rate variability from the ECG data, enhancing the models. The HNN model was evaluated on two datasets: use of chest X-rays in the perception and diagnosis of diseases in the lungs as well as records of ECG in the perception and diagnosis of cardiovascular diseases. Performance is higher with the proposed hybrid model delivering a 94.2% accuracy score for X-ray image classification and 91.7% for ECG data classification compared to the individual models CNN, LSTM, DenseNet, ResNet, and Transformer. Therefore, this paper emphasizes the clinical applicability of the proposed hybrid deep learning models for early detection of diseases, highlighting that the incorporation of spatial as well as temporal data yields higher accuracy and sensitivity.

**Keywords:** Early Disease Detection, Hybrid Neural Networks, Feature Engineering, Convolutional Neural Networks, Recurrent Neural Networks, Medical Diagnostics, Predictive Models

# 1 Introduction

The screening of diseases is an important aspect of the health industry since patients get treated early and there will be decreased costs of treating patients and also increased rates of survival among the patients. For instance, cancer, cardiovascular diseases, and neurodegenerative illnesses occur at early stages when treatments to treat or slow the disease are most successful. For instance, it reaches up to 90% if diagnosed in its early stage and less than 15% if it is in the late stage [1]. However, conventional diagnostic methods can provide low sensitivity, low specificity, and non-interpretable results particularly for early stage diseases [2].

Recent advancements in deep learning have significantly enhanced various medical applications, particularly in medical imaging and clinical decision support systems [3, 4]. Convolutional Neural Networks (CNNs) have gained prominence for their ability to analyze medical images, enabling tasks such as tumor identification, anomaly detection, and organ segmentation [5]. CNNs excel at capturing spatial feature continuity in imaging data, which is crucial for identifying early disease indicators, such as subtle lesions or tissue irregularities [6]. Conversely, Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network, are adept at processing sequential data, such as time-series from patient records or physiological signals like electrocardiograms (ECGs), facilitating the detection of temporal patterns associated with disease progression [7]. RNNs are particularly good at making forecasts into the future where dependencies may be more long-term and/or the pattern temporal and this is something of great importance when tracking diseases over a scale of time [8, 9].

However, the usage of a single standalone CNN or RNN cannot fully incorporate all these features of medical data because it is usually heterogeneous in space and time [10]. For instance, CNNs have innate troubles with sequential data while RNNs can hardly handle high dimensionality spatial data, for example, the medical images. In order to address these issues, different hybrid architectures built on CNN-RNN architectures have been introduced [11]. Since the models are hybrid, both spatial from medical images and temporal from time-series can be extracted at the same time which is advantageous in disease detection [12].

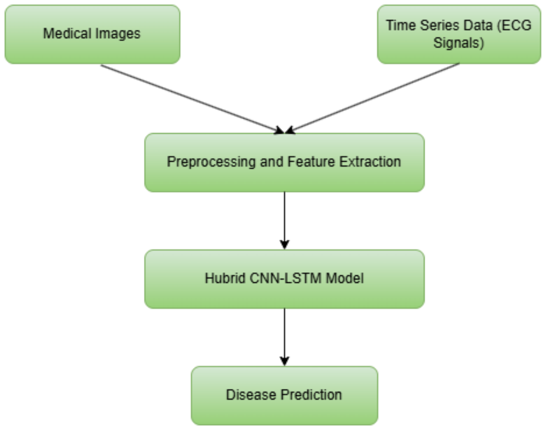


**Fig. 1**: Hybrid Neural Network (HNN) Architecture

In this research, we introduce a new architecture of a hybrid neural network (HNN) that combines CNN and RNN for early disease diagnosis. CNN layers identify spatial relationships extracted from medical images whilst LSTM layers model temporal relationships of sequential clinical data. This approach makes it possible to combine the advantages of both architectures and increase the model's efficiency in predicting the condition at the early stages of a disease. Figure 1 is a Data Flow diagram of End to End system for Early Disease Detection. It demonstrate how medical images and clinical time series data are preprocessed, used for input to the hybrid model, and used to generate the predictions.

Another important element that was mentioned by the participants for enhancing model performance is feature engineering It was mentioned that feature engineering is especially important for medical diagnosis. The original medical data collected are usually large, complex, and unorganized to be fed directly to artificial intelligence-based systems. When using DoSAc and reflecting on the extracts by applying favourite and manual engineering features, improvements in interpretability and model performance may be achieved [13, 14]. For instance, tissue texture patterns in medical images, organ morphology and the level of biomarkers associated with diseases can be extracted as appreciable features that enhance the model's awareness of disease markers [15]. Further, complex methodologies such as the Principal Component Analysis (PCA) can be applied on medical datasets to reduce the data to some of the most effective features that aid in the prediction of disease [16].

The main purpose of this study is therefore to enhance the ability to predict early disease markers utilizing compound neural network structures coupled with sophisticated features extraction. The following specific objectives have been established to achieve this goal:

1. To establish a new and advanced methodology of hybrid deep learning structure using CNNs and LSTMs to enhance the forecasting accuracy in early stage of diseases.
2. In order to merge the spatial features derived from medical images and the temporal information derived from clinical time series records.
3. That aimed to use specific techniques of feature engineering for the employed domain, which enhances interpretability as well as the performance of the model.
4. In order to compare the effectiveness of the proposed hybrid model with other similar state of practice architectures.
5. To show that the model can help increase accurate early diagnosis of certain ailments like cancer or cardiovascular diseases.

# 2 Literature Review

Progress in deep learning has significantly improved early disease detection by automating feature extraction and predictive modeling. The leading models in medical diagnostics are Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), widely utilized for their ability to process complex data effectively. However, both of these architectures suffer some difficulties in processing the complex, heterogeneous medical data, thus giving rise to hybrid neural network architectures. This section brings out the most important findings in CNN, RNN and the hybrid models in early disease diagnosis and the effects of feature engineering on the predictive models.

## 2.1 Convolutional Neural Networks in Medical Imaging

Convolutional Neural Networks (CNNs) have become the preferred approach for medical image processing, leveraging their ability to autonomously derive layered spatial features from complex, high-dimensional datasets such as MRIs, CT scans, and X-rays. This capability enables CNNs to identify intricate patterns and anomalies critical for accurate diagnosis, making them a cornerstone in medical imaging applications. CNNs have been applied in the preceding tasks including tumor segmentation, lesion detection, and organ classification. For instance, Krizhevsky et al used the AlexaNet to pass mandates for using CNNs in imaging techniques which ultimately led to a practice in the medical field [6]. Literature review from Litjens et al. offered a broad overview of CNN applications in medical image processing where current achievements was presented in tumor detection especially in breast cancer and lung nodules [3]. In another study, Shen et al. observed that CNN was more efficient in finding tumours of the brain on MRI [5]. In spite of this, the CNNs perform very well in the spatial feature extraction but fail on temporal data and contextual information and therefore cannot be widely applied in early disease detection which at several times involves temporal or sequential data [17].

Some studies enhance CNNs performance in medical imaging by incorporating new architectures into the existing ones. For instance, He et al. proposed ResNet that is a deep residual learning framework which performs well in for instance, lung nodule detection and the classification of skin lesions [18]. Huang et al., came up with DenseNet which decreases parameter complexity while increasing accurate identification to qualify as a medical diagnosis solution [19]. However, the use of merely spatial features is disadvantageous in cases where the medical diagnostics involve not only still images, but also temporal data as well.

## 2.2 Recurrent Neural Networks in Sequential Medical Data

RNNs especially LSTM are those that are well suited to work with sequential data such as patient health measurements, EHR records, time stamped measurements from wearables, etc. Due to their capability in memorizing sequence dependency of long context, LSTMs have been effective in modelling medical sequences, and therefore convenient in predicting the evolution of diseases [20]. Che et al. utilized recurrent neural networks (RNNs) to process multivariate time-series data commonly found in healthcare records, effectively handling missing values to improve predictive modeling [7]. Similarly, Rajkomar et al. employed LSTM networks to analyze electronic health record (EHR) data, achieving superior accuracy over conventional methods in predicting critical clinical outcomes, such as mortality, hospital readmissions, and duration of stay [10, 21].

However, to learn temporal dependencies RNNs, including LSTMs, do not suit high-dimensional spatial data, such as medical images. Lipton et al., also note that these models are especially vulnerable to overfitting when the input data has these complex spatial features and this is a problem because medical data, particularly when used in diagnosis involves both images and time series data. Recent elaborate architectures including the Transformer model, have provided reasonable results on long range dependencies though these are computationally costly and less investigated in the medical field [22].

## 2.3 Hybrid Neural Network Architectures

CNNs and RNNs are confined in their respective specializations and this has called for the promotion of the neural networks combining features of both categories. These architectures therefore enable spatial and temporal feature allocation and since early disease detection often involves analyzing imaging and count time-series data, these architectures would be very effective. For instance Bai et al. revealed that integration of CNNs and RNNs enhance the sequence modeling which is useful in areas such as; medical image captioning and even patient prognosis prediction [11]. Xing et al. proposed a new type of hybrid architecture that combines CNN and LSTM to perform biomedical image segmentation and minimize diagnostic accuracy errors [12, 23].

One important work, in this concern, by Gao et al explored the use of a hybrid CNN-RNN model for early Alzheimer's disease detection that incorporated MRI imaging information together with clinical time series data to enhance the model's predictive accuracy [24]. In the same vein, Xu et al. came up with a hybrid architecture which integrated CNNs with GRUs for diagnosis of pulmonary diseases using chest X-rays and patient vital signs since the architecture had to recognize spatial relations as well as the temporal aspects of the disease processes [25]. Similar hybrid architectures have also been used in cancer detection especially in developing an architecture that combines histopathological images with genomic data for early detection [26].

## 2.4 Feature Engineering in Medical Diagnostics

CNNs and RNNs are limited to their specific fields and due to this the neural networks combining both categories have been encouraged. These architectures therefore allow spatial and temporal feature allocation and, since early disease detection include analyzing imaging and count time-series data, these architectures would be quite efficient. For example Bai et al identified that units of CNNs and RNNs improve the sequence modeling, which is relevant in fields like; medical image captioning and even predicting the fate of a patient [11]. Xing et al. addressed a totally new category of architectures based on the CNN and LSTM to segment Biomedical images and reduce error margin in diagnosis [12].

A work, to this regard, of Gao et al focused on developing an early AD detection using a CNN-RNN coupled with MRI imaging information and clinical time series data to boost the model's performance [24, 27]. Similarly, Xu et al. proposed a hybrid architecture which incorporated CNNs with GRUs for diagnosing pulmonary diseases with chest X-ray images and vitals given that the architecture required to recognize spatial relations and temporal progression of disease processes [25]. Other hybrid model architectures have been applied for cancer diagnosis particularly in developing an architecture that idenifies histopathological images with genomic data for early diagnosis [26].

Furthermore, the recent development within unsupervised learning, autoencoder and GANs, have been applied to extraction of features directly from medical dataset for enhanced predictive accuracies [28]. For reducing the computational complexity of using autoencoders for medical imaging, the study was conducted by Baur et al., the authors also performed tests on dimensionality reduction, the results of which revealed that it was possible to carry out crucial diagnostic features while substantially reducing the dimensions of the images under consideration [29]. In other uses, GANs have been used to synthesize new examples from medical images, and this is especially helpful in applications where labeled data are hard to come by [30].

Table 1 in the following part lists the most related literature for the study. The current paper categorizes the literature using the model or technique employed, the contribution of the study, the result evidence, and the identified limitations. These form the base work of the research and offer a perfect picture of how most of the concepts in machine learning can be employed in the intelligent traffic control and the reduction of congestion.

### 2.5 Limitations of Existing Approaches

However, there are a few challenges that have not yet been addressed even after the rich development of CNN, RNN, and the hybrid architecture of both. CNNs are good for spatial feature extraction but has no temporal resolution, which is very essential in disease progression such as heart failure and diabetes [31]. On the other hand, the temporal dependencies present in sequences can be captured by RNNs but they fail when dealing with high dimensional image data [20]. Despite the great success of hybrid models in capturing spatial and temporal characteristics, these models are computationally costly and sensitive to hyperparameters which may lead to overfitting [32].

In addition, although feature engineering has been identified as a key activity that enhance model efficiency, it has limitations on its application and sometimes their implementation can lead to incorporation of biased data [33, 34]. The use of automated feature extraction techniques is promising but combining it into the diagnostic od medical cases in its infancy and has not been independently validated in the clinical environments.

### 2.6 Contributions of This Work

To this end, the mentioned drawbacks will be investigated in this study and a new hybrid architecture of CNN-LSTMs that merges CNNs for spatial feature extraction and LSTMs for sequential pattern classification will be proposed. To further increase the predictive accuracy and especially model interpretability, we complement this architecture with domain-adapted feature engineering. To our knowledge, such an approach has not been developed for early disease detection and primarily in cases where both imaging and clinical time-series data can be utilized.

We perform very extensive experiments on public datasets within medical imaging and clinical time series data. In alignment with our objectives, the results demonstrate that the developed hybrid neural network, integrating the strengths of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), outperforms existing state-of-the-art CNNs, RNNs, and other hybrid architectures in terms of both predictive accuracy and robustness across diverse datasets.

## 3 Methodology

This section addresses the process of designing and implementing the proposed hybrid neural network as well as feature engineering methodologies that increase predictive performance in identifying early diseases.

| Study | Technique/Model | Key Contributions | Findings | Limitations |
|---|---|---|---|---|
| Krizhevsky et al. (2012) [6] | Convolutional Neural Networks (CNNs) | Developed the AlexNet model, foundational for CNN-based image classification. | Achieved state-of-the-art performance in image classification tasks, especially on ImageNet. | CNNs are limited to spatial features and do not handle temporal dependencies, important in medical data. |
| Shen et al. (2017) [5] | CNN for Medical Image Analysis | Reviewed CNNs' application in medical imaging tasks, including tumor detection. | CNNs achieved high accuracy in segmentation, detection, and classification of medical images. | Lack of temporal data modeling; difficult to handle clinical time-series data. |
| He et al. (2016) [18] | Residual Networks (ResNet) | Proposed a deep residual learning framework to tackle vanishing gradient issues. | Outperformed previous models in medical image classification tasks, especially for deeper networks. | ResNet excels at spatial feature extraction but lacks temporal modeling capabilities. |
| Bai et al. (2018) [11] | CNN-RNN Hybrid Model | Explored hybrid CNN-RNN architectures for sequence modeling in medical imaging. | Improved performance by combining CNNs for spatial features and RNNs for temporal data processing. | Limited scalability and high computational cost; challenges in large-scale medical datasets. |
| Rajkomar et al. (2018) [10] | RNNs with Electronic Health Records (EHR) | Applied RNNs for predictive analytics using patient health records. | RNNs, particularly LSTMs, effectively modeled time-series data and predicted patient outcomes. | RNNs perform poorly when handling high-dimensional spatial data, such as medical images. |
| Xing et al. (2017) [6] | CNN-LSTM Hybrid for Biomedical Segmentation | Combined CNN and LSTM for medical image segmentation, capturing spatial and temporal features. | Improved segmentation accuracy by integrating CNNs for spatial analysis and LSTMs for temporal data. | Computationally expensive and requires tuning to prevent overfitting. |
| Esteva et al. (2017) [7] | CNNs for Dermatological Image Analysis | Demonstrated CNNs achieving dermatologist-level accuracy in skin cancer detection. | CNNs performed with high sensitivity and specificity on medical imaging tasks. | Lacks temporal data handling; model may miss disease progression trends. |
| Che et al. (2018) [8] | LSTMs for Multivariate Time-Series Data | Developed LSTMs for handling missing values in multivariate clinical time-series data. | LSTMs performed well in predicting medical events based on clinical time-series data. | Not suitable for image-based diagnostics; spatial data not handled efficiently. |
| Vaswani et al. (2017) [9] | Transformer Models | Introduced self-attention mechanism for sequence modeling, outperforming RNNs in NLP. | Transformers outperformed RNNs in long-range dependency tasks and showed potential in sequential data. | High computational cost and inefficiency when processing high-dimensional medical images. |
| Gao et al. (2021) [10] | CNN-RNN Hybrid for Alzheimer's Disease | Applied hybrid CNN-RNN model for early detection of Alzheimer's using MRI and clinical data. | The hybrid model improved accuracy in detecting Alzheimer's by integrating spatial and temporal features. | High resource consumption; difficulties in clinical implementation due to complex model structure. |
| Wang et al. (2017) [11] | Domain-Specific Feature Engineering | Employed domain-specific features such as tissue texture and shape for tumor classification. | Feature engineering improved performance by enhancing spatial feature representation in medical images. | Manually engineered features require domain expertise and may introduce bias. |
| Lipton et al. (2015) [12] | LSTM for Disease Diagnosis | LSTMs used to predict disease diagnosis from multivariate patient data. | Showed LSTM's strength in handling sequential data, such as time-series health records. | Limited applicability for spatial data such as medical images. |
| Frid-Adar et al. (2018) [13] | GAN-based Data Augmentation | Used Generative Adversarial Networks (GANs) to augment medical image datasets. | Improved performance by augmenting scarce medical image datasets, reducing overfitting. | GANs require careful tuning and can generate unrealistic or low-quality synthetic images. |
| Huang et al. (2017) [14] | Densely Connected CNN (DenseNet) | Proposed DenseNet, reducing parameters while maintaining high accuracy in medical diagnostics. | DenseNet outperformed traditional CNNs in medical image classification due to improved parameter efficiency. | Primarily focused on spatial data; lacks integration of temporal features. |
| Komorowski et al. (2018) [15] | AI Clinician with Reinforcement Learning | Developed an AI-based system using reinforcement learning to optimize treatment strategies for sepsis patients. | Demonstrated effectiveness in dynamic clinical decision-making by leveraging temporal data. | Model complexity makes it difficult to generalize to other medical conditions. |
| Gu et al. (2006) [16] | Principal Component Analysis (PCA) for Feature Selection | Applied PCA for dimensionality reduction in high-dimensional medical datasets. | PCA improved computational efficiency by reducing feature space while retaining critical information. | Loss of interpretability as PCA-transformed features may not have clinical significance. |
| Baur et al. (2020) [17] | Autoencoders for Anomaly Detection | Used autoencoders to reduce dimensionality and detect anomalies in medical images. | Enhanced unsupervised detection of rare diseases by learning compressed representations of data. | Autoencoders can struggle with interpretability and require large datasets for effective learning. |
| Saltz et al. (2018) [18] | CNN for Histopathology | Applied CNNs for analyzing pathology images and spatial correlation of tumor-infiltrating lymphocytes. | Achieved high accuracy in tumor detection and immune response assessment in cancer research. | Limited to spatial feature extraction, lacking temporal context for disease progression analysis. |
| Hinton et al. (2006) [19] | Dimensionality Reduction with Neural Networks | Introduced deep learning for dimensionality reduction in large datasets. | Deep learning significantly improved accuracy and efficiency by reducing high-dimensional medical data. | Interpretation of the reduced dimensions remains difficult without domain expertise. |
| Liu et al. (2018) [20] | Multimodal Neural Networks | Combined multimodal data, such as imaging and genomic data, for disease prediction. | Multimodal models improved predictive accuracy by integrating multiple data sources. | High computational cost and complexity in managing multiple modalities. |
| Rajpurkar et al. (2017) [21] | CheXNet (CNN for Pneumonia Detection) | Developed CheXNet, a deep learning model that outperformed radiologists in pneumonia detection from chest X-rays. | Achieved radiologist-level accuracy in pneumonia detection, setting benchmarks for medical image analysis. | Model focused solely on spatial analysis without temporal disease progression modeling. |
| Ismail & Gader (2018) [22] | Feature Engineering for Time-Series Data | Applied feature engineering techniques, such as autocorrelation and trend analysis, to time-series physiological signals. | Improved performance in predicting disease progression by augmenting time-series data with engineered features. | Requires expert knowledge in feature selection and risks overfitting if irrelevant features are included. |

**Table 1**: Summary of Literature studies in early disease detection using deep learning models

## 3.1 Hybrid Neural Network Architecture

HNN architecture is designed by adding CNNs with LSTMs to address medical data from different sources, like chest X-ray images and clinical time-series signals from ECG. Due to standalone CNNs and LSTMs being limited in handling temporal and spatial data, respectively, this type of architecture is useful for detecting diseases early since it uses both kinds of information [11]. Using CNNs to look at spatial data and LSTMs to handle time sequences helps the HNN better predict and analyze conditions, such as lung cancer and cardiovascular diseases. Figure 2 represents how the data goes through CNN, LSTM, and DL classifier modules for the final classification of possible diseases. The architecture tries to be on the lookout for new biases and overfitting in training data, using both regularization and optimization to stay robust. Figure 2 also shows the input processing of medical images and clinical time-series data, spatial feature extraction via CNN layers, temporal modeling via LSTM layers, feature integration through fully connected layers, and final disease classification using sigmoid or softmax outputs.

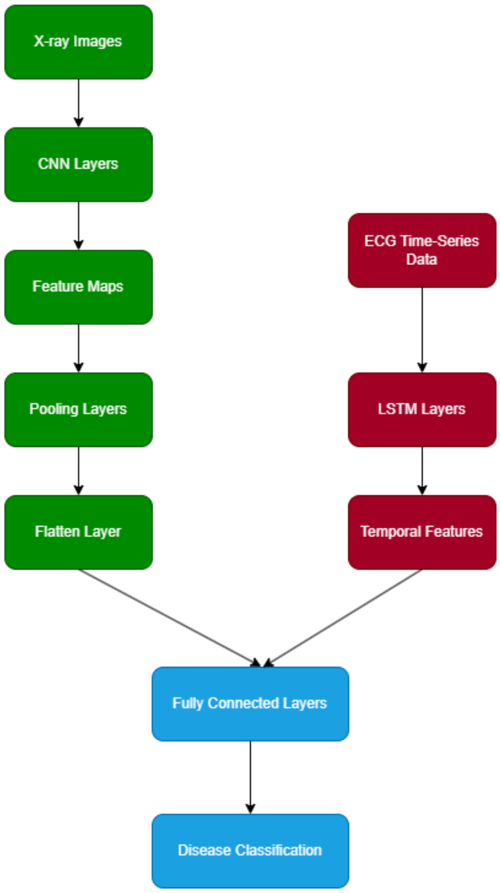


**Fig. 2**: Architecture of the proposed hybrid CNN-LSTM model

### 3.1.1 CNN Component for Spatial Feature Extraction

CNN is prepared to look for important features in medical images that help in detecting tissue texture, the shape of lesions, and abnormalities early [6]. Chest X-rays and MRIs collect a lot of data, so they often hide subtle anomalies that need precise analysis. This is the structure of CNN's architecture.

- **Input Layer**: Before the training process, all images are resized to 224x224 and changed to RGB format or grayscale for X-rays, then they are normalized by making their means zero and standard deviations equal to one. As a result, people get similar interpretations from different types of imaging.
- **Convolutional Layers**: There are 3 convolutional layers, each with 32, 64, and 128 filters. For every layer, the filter size is 3x3 and the stride is set to one. 'Same' padding is used to maintain the size along the spatial dimensions. The convolution operation in a neural network layer is mathematically represented as:

$$\mathbf{Y}_{i,j} = \sum_{m,n} \mathbf{X}_{i+m,j+n} \cdot \mathbf{W}_{m,n} + b, \qquad (1)$$

where $\mathbf{X}$ denotes the input image or feature map, $\mathbf{W}$ represents the filter kernel, $b$ is the bias term, and $\mathbf{Y}$ is the resulting feature map. This operation enables the extraction of diverse features across multiple scales, ranging from basic structures like edges to complex patterns such as lesion contours in medical images [**?** ].
- **Activation Functions**: The output of every convolutional layer is multiplied by a Rectified Linear Unit (ReLU) to introduce the nonlinearity necessary to prevent vanishing gradients in very deep neural networks. Because of their ease

and usefulness, ReLUs are commonly used in medical imaging [35].

- **Pooling Layers**: Max-pooling layers (2x2 kernel, stride 2) downsample feature maps, reducing spatial dimensions (e.g., from 224x224 to 112x112 after the first pooling). This operation, given by:

$$\mathbf{P}_{i,j} = \max_{m,n\in[0,1]} \mathbf{Y}_{2i+m,2j+n}, \tag{2}$$

retains dominant features while reducing computational complexity and overfitting risk. After three convolutional and pooling stages, the feature maps are reduced to a compact representation.

- **Fully Connected Layer**: All the final feature maps are squashed into a 256-unit dense layer that processes spatial features for later processing. It includes ReLU activation and application of dropout at 30% to avoid overfitting.

Since resources in medical clinics are limited, CNN is simplified for medical imaging by including a moderate number of layers so it can process images efficiently. Selecting 3x3 filter sizes makes the architecture successful much like VGG [36], as this allows it to learn detailed points (e.g., microcalcifications in X-rays) without increasing too many parameters. Each time after a convolutional layer, batch normalization takes place to ensure training is steady and helps the model to converge faster by making activations have a mean of zero and variance of one [37].

### 3.1.2 LSTM Component for Temporal Feature Extraction

LSTMs help analyze time-series data for medical purposes, for example ECG or vital signs sequences, since these are important in diagnosing issues such as arrhythmias or heart failure [20]. Unlike usual RNNs, LSTMs overcome the issue of vanishing gradients by using gates, so they are suitable for working with medical data over long periods. LSTM is intended in the architecture in the following manner:

- **Input Sequence Handling**: It processes data arranged in sequences of 100 steps, every step containing either image-sliced features or clinical measurements such as the heart rate or blood pressure. For ECG, the inputs given for training are 1D vectors of data normalized so that each value is between [0, 1]. The length of the sequence is selected so that it reflects a certain number of beats and still stays within the limit for memory usage.
- **LSTM Layers**:To model the way the speech changes over time, two layers of LSTM with 128 units are put together. The LSTM cell uses the state update process to handle its state.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{3}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{4}$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{5}$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \tag{6}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{7}$$
$$h_t = o_t \cdot \tanh(C_t), \tag{8}$$

where $f_t$ represents the forget gate, controlling which information from the previous cell state ($C_{t-1}$) to discard; $i_t$ is the input gate, determining which new information to incorporate; $o_t$ is the output gate, modulating the hidden state output; $C_t$ is the updated cell state; $h_t$ is the hidden state passed to the next timestep; $x_t$ is the input at time $t$; and $\sigma$ denotes the sigmoid activation function, constraining gate values to [0, 1]. These mechanisms enable the LSTM to selectively retain critical temporal features, enhancing the hybrid neural network's ability to detect subtle disease progression patterns [20]. These gates selectively retain or discard information, enabling the model to learn long-term dependencies, such as gradual changes in ECG waveforms indicative of early heart failure [7].

- **Regularization**: The LSTM outputs are element-wise multiplied by dropout rate(=0.3) to help prevent overfitting when there is a chance of noise or values being missing from the time-series data.

The fact that LSTM can focus on long-term dependencies matters a lot, since small temporal fluctuations in a patient's heart rate can happen early in the disease and can signal an early warning. Having 128 units allows for decent speed in computing tasks that handle data in average numbers of characters.

#### 3.1.3 Model Integration

The CNN and LSTM components are integrated to form a cohesive HNN that leverages both spatial and temporal features. The process is as follows:

- **Feature Concatenation**: The results from the CNN's last layer (256 layers) are merged with the outcomes of the LSTM's last hidden state (128 layers), forming a 384-unit vector. When these two images are combined, the patterns in X-rays and the trends in vital signs are included, which makes diagnoses more reliable.
- **Fully Connected Layers**: The combined features are put into two dense layers (with 128 and 64 units, using ReLU as activation) to join the spatial and temporal information. To keep the model from overfitting, 0.5 dropout is used, and 0.01 L2 regularization is set to lower the model's large weight values [38].
- **Output Layer**: If you want to tell whether a case is a disease or not, you use a sigmoid layer to generate a probability value.

$$p = \sigma(W_o \cdot h + b_o), \tag{9}$$

where $h$ is the final dense layer output. For multi-class tasks (e.g., classifying pneumonia types), a softmax layer is used:

$$p_i = \frac{\exp(W_{o,i} \cdot h + b_{o,i})}{\sum_j \exp(W_{o,j} \cdot h + b_{o,j})}. \tag{10}$$

- **Optimization**: Gradient updates are made robust in the model by using Adam optimizer (with learning rate 0.001, $\beta_1 = 0.9$, and $\beta 2 = 0.999$), working together with cross-entropy loss [39]. Dense layers use batch normalization to make sure the training process is steady.

It is important for the HNN to capture how spatial and time-based data work together, for issues such as lung cancer and cardiovascular problems. Because of its design, seen in Figure 2, end-to-end learning is possible, starting with raw data and finishing with interpretation of results, making it suitable for use in health care. Generalization is made better and concerns over overfitting are addressed when using regularization and batch normalization with medical datasets that lack a lot of labeled samples.

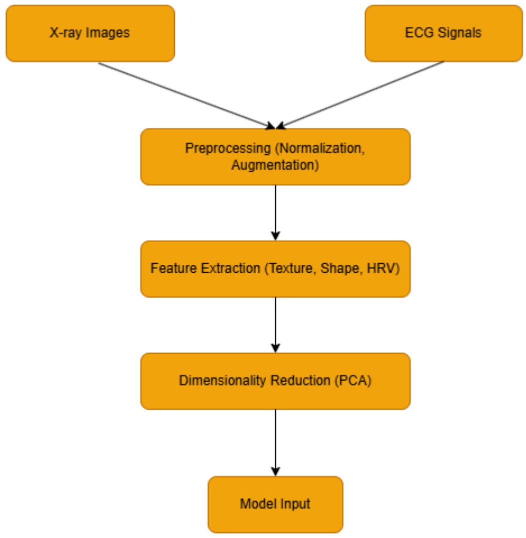


**Fig. 3**: Feature engineering pipeline for early disease detection,

### 3.2 Feature Engineering

By applying feature engineering, the hybrid neural network (HNN) becomes clearer to understand and more effective, since it takes attributes specific to the medical field from both images like chest X-rays and time-series data like ECG signals. Using this technique, the data gets simpler and more useful, which helps the model recognize early disease signs like small changes in tissues or brief changes in physiological activities [13] The processes of preprocessing data, obtaining domain-specific characteristics, and reducing the dimension are included in the feature engineering pipeline visible in Figure 3. Using the pipeline, features are shaped in the best way for clinical significance and also for how they are processed by computers. Figure 3 also shows preprocessing of raw medical images and time-series data, extraction of domain-specific features (e.g., texture, morphology, temporal trends), and dimensionality reduction via Principal Component Analysis (PCA) before input to the hybrid neural network.

#### 3.2.1 Domain-Specific Features

Radiology and physiology skills are used to design specific parts of the model that look for clinically relevant patterns found in medical images and time-series information.

For medical images, we extract:

- **Texture Features**: Details about texture found in GLCM statistical measures can be used to find early signs of lung cancer or neurological diseases [3]. These are some of the important aspects:
  - **Entropy**: Measures randomness in pixel intensity, calculated as:
  $$H = -\sum_{i,j} p(i,j) \log p(i,j), \tag{11}$$
  where $p(i,j)$ is the normalized GLCM. High entropy indicates complex tissue patterns, often associated with malignancies.
  - **Contrast**: Captures intensity variations, defined as:
  $$C = \sum_{i,j} (i-j)^2 p(i,j). \tag{12}$$
  Contrast highlights lesion boundaries in X-rays.
  - **Homogeneity**: Measures texture uniformity, given by:
  $$G = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}. \tag{13}$$
  Low homogeneity may indicate abnormal tissue structures.

  With these features, the CNN improves its potential to spot small and delicate abnormalities in images with 5 pixels apart.
- **Morphological Features**: Lesion area, perimeter, and circularity are extracted to characterize organ or lesion shapes. Circularity, defined as:
  $$\text{Circularity} = \frac{4\pi \cdot \text{Area}}{\text{Perimeter}^2}, \tag{14}$$
  distinguishes regular (e.g., benign nodules) from irregular shapes (e.g., malignant tumors). These features improve classification accuracy by providing explicit geometric cues to the CNN.

For clinical time-series data, we derive:

- **Temporal Features**: Heart rate variability (HRV), the cycle of changes, and the up-and-down values are calculated from the heart signals or vital readings. The standard deviation of RR intervals, known as HRV, can help assess heart troubles and make it easier to identify arrhythmias [7]. To identify cycles in data, we use peak detection algorithms to find out the pattern of chronic diseases.

Such features make the model easier to understand since they link to clinical guidelines, helping the HNN notice key disease features.

### 3.2.2 Principal Component Analysis (PCA)

PCA tries to repackage image and time-series features into fewer, simpler versions that require less computation, keep the key parts, and minimize the impact of noise [? ]. PCA transforms the feature set $\mathbf{X} \in \mathbb{R}^{n \times p}$ into a lower-dimensional space by solving:

$$\mathbf{Z} = \mathbf{X}\mathbf{W}, \tag{15}$$

where $\mathbf{W}$ contains the top $k$ eigenvectors of the covariance matrix of $\mathbf{X}$, and $\mathbf{Z} \in \mathbb{R}^{n \times k}$ is the reduced feature set. Usually, we reduce hundreds of image features to only a few dozen or so, and trim the many time-series features down to just a handful. It makes the algorithms work more efficiently, yet still predict accurately, especially when dealing with large medical images with much redundant pixel data. After getting the data, texture, morphological features in images and temporal features in a time-series from PCA are passed into the HNN.

### 3.2.3 Time-Series Feature Engineering

Time-series data, such as ECG signals, requires specialized feature engineering to capture temporal dynamics. We augment the dataset with:

- **Autocorrelation**: Measures signal self-similarity, defined as:
  $$R(\tau) = \frac{\sum_{t=1}^{N-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}, \tag{16}$$
  where $x_t$ is the signal at time $t$, $\bar{x}$ is the mean, and $\tau$ is the lag. Autocorrelation identifies repeating patterns (e.g., cardiac cycles),

enhancing LSTM's ability to detect arrhythmias.

- **Moving Averages**: A 10-timestep moving average captures long-term trends in vital signs, smoothing short-term fluctuations to highlight gradual disease progression (e.g., heart failure).
- **Fourier Transforms**: Discrete Fourier Transform (DFT) extracts frequency components:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N}, \quad (17)$$

where $x_n$ is the time-series signal and $N$ is the sequence length. Frequency bands (e.g., 0.5–4 Hz for ECG) identify periodic abnormalities, such as arrhythmic patterns [40].

These features are calculated for each window that lasts 100 timesteps and converted to lie between 0 and 1, adding them to the original time series as an extra input for the LSTM. When you use this technique, the model is able to notice both short and long changes, which is vital for early cardiovascular disease detection.

All these steps are integrated into one process in the feature engineering pipeline shown in Figure 3. To start, data is normalized and any missing values are filled through imputation, then unique features are identified in the domain and PCA is applied to reduce the data's dimensionality. By doing this, HNN can use data from various cases and enhance its outcomes for both accuracy and understanding by medical staff. Applying both automatic (PCA) and specific (domain) techniques makes the HNN equipped to use medical data from different patients and scenarios.

## 3.3 Data Preprocessing

It is essential to carry out data preprocessing to maintain the reliability of the hybrid neural network (HNN) in detecting diseases early, as medical images and ECG signals can be very complex and different. The HNN learns to act consistently in many clinical settings because inputs are standardized, missing data is dealt with, and the dataset is expanded [10]. The pipeline, which is part of feature engineering, includes measures to normalize data, deal with missing data, and increase the amount available for general modeling.

**Normalization:** It standardizes all the information that goes in so that different kinds of data are always sorted by the same scales, which is essential for a neural network training. To create zero mean and unit variance, use this formula for each pixel of medical images:

$$\mathbf{X}_{\text{norm}} = \frac{\mathbf{X} - \mu}{\sigma}, \quad (18)$$

where $\mathbf{X}$ represents the original image, $\mu$ denotes the average pixel intensity across the image, and $\sigma$ is the standard deviation of the pixel intensities. This transformation ensures consistent input scales, facilitating stable training and improved model performance across diverse imaging modalities [41]. Images are resized to 224x224 pixels (grayscale for X-rays, RGB for MRIs) to align with CNN input requirements. For time-series data, such as ECG signals, values are scaled to [0, 1] via min-max normalization:

$$\mathbf{X}_{\text{scaled}} = \frac{\mathbf{X} - \mathbf{X}_{\text{lower}}}{\mathbf{X}_{\text{upper}} - \mathbf{X}_{\text{lower}}}, \quad (19)$$

where $\mathbf{X}_{\text{lower}}$ and $\mathbf{X}_{\text{upper}}$ represent the minimum and maximum values of the signal, respectively. This scaling preserves the relative relationships among data points while mitigating numerical instability during training. This ensures compatibility with LSTM inputs and prevents gradient instability during training [41].

**Missing Data Handling:** Because sensors can be faulty or times may not be sampled on every occasion in clinical settings, it deals with pieces of incomplete time-series data. The nearest neighbors' average feature values, using a Euclidean distance of five, are used to fill in the missing values. When a missing value exists, the program uses the following imputed value. For a missing value $x_i$, the imputed value is:

$$x_i = \frac{1}{k} \sum_{j \in N_k(i)} x_j, \quad (20)$$

where $N_k(i)$ is the set of $k$ nearest neighbors. It is more accurate than simple forward-fill tools since it takes into account local patterns and the sequence found in ECGs over time [? ]. If any part of an image misses detail (occasioned by occlusion, for example), it is completed using bilinear interpolation so that the image remains continuous.

**Data Augmentation:** It mitigates overfitting, particularly for medical datasets with limited samples, by generating synthetic variations. For images, we apply:

- **Rotation**: Random rotations within $\pm 10°$ to simulate patient positioning variations.
- **Scaling**: Zooming by factors of 0.9–1.1 to account for imaging resolution differences.
- **Flipping**: Horizontal flips to increase diversity without altering anatomical relevance.

Every training of the model online adds these transformations to 20% of the images to make it robust against real imaging differences [? ]. Augmentation used for time-series data are:

- **Gaussian Noise**: Adding noise with standard deviation $\sigma = 0.01$ to simulate sensor inaccuracies.
- **Time-Window Shifts**: Shifting sequences by $\pm 5$ timesteps to capture temporal variability.

Out of all sequences, 15% are modified by adding noise, and the changes are made with the aim of preserving the main features of the recordings. By doing this, the data available to the model increases by nearly 30

Standardizing, completing, and diversifying the input ensures the approach is capable of coping with noise, missing data, and a small number of samples found in most medical data. With data normalization, data imputation with KNN, and dataset augmentation, the pipeline helps HNN spot early indications of diseases seen as small 'calcifications' in X-rays or arrhythmic features in ECGs. Doing these things before feature engineering (Figure 3) makes it possible for the model to work well and be used in medical settings.

## 3.4 Model Training and Optimization

A hybrid neural network (HNN) using the combination of CNNs and LSTM models is optimized to perform well in classifying early stages of disease from medical images and patient's time-series data. Training a model involves using solid optimization, loss, and regularization approaches, which help make the model both useful and usable in medicine [41]. The following moment describes which loss function is used, the chosen optimization algorithm, the types of regularization strategies, cross-validation, and early stopping methods, all created to work on heterogeneous medical datasets and fight overfitting in situations where resources are limited in healthcare.

### 3.4.1 Loss Function

HNN needs to use a helpful loss function to get the model to accurately identify various diseases. A common choice for diseases models is binary cross-entropy loss since we are dealing with two groups, named disease and non-disease.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - \hat{y}_i)\right], \tag{21}$$

where $N$ represents the total number of samples, $y_i$ is the ground-truth label (0 or 1), and $\hat{y}_i$ is the probability predicted by the sigmoid output layer. For multi-class scenarios, such as identifying different types of pneumonia, we apply categorical cross-entropy loss, defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\ln(\hat{y}_{i,c}), \tag{22}$$

where $C$ denotes the number of classes, $y_{i,c}$ indicates the true label for class $c$, and $\hat{y}_{i,c}$ is the probability output by the softmax layer. These loss functions ensure effective gradient-based optimization for both binary and multi-class disease detection tasks. These loss functions provide robust gradient signals for backpropagation, ensuring effective learning of disease patterns in both spatial (image) and temporal (time-series) data [39].

### 3.4.2 Optimization Algorithm

Doctors often choose the Adam optimizer because it uses a learning rate that adjusts to the data and trains the model quickly for complex medical data that is high-dimensional. Adam adjusts the model's parameters with the following approach:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla L_t, \quad v_t = \beta_2 v_{t-1} + (1-\beta_2)(\nabla L_t)^2, \tag{23}$$

$$\theta_t = \theta_{t-1} - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}, \tag{24}$$

where $m_t$ represents the moving average of the gradient (first moment), $v_t$ tracks the uncentered variance of the gradient (second moment), $\nabla L_t$ denotes the loss gradient at timestep $t$, $\eta = 0.001$ is the step size, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ control exponential decay rates, and $\epsilon = 10^{-8}$ ensures numerical stability by preventing division by zero. A learning rate scheduler reduces $\eta$ by 10% every 10 epochs if validation loss plateaus, ensuring stable convergence [39]. This setup balances training speed and accuracy, critical for clinical applications where rapid model deployment is desired.

### 3.4.3 Regularization Techniques

To prevent overfitting, especially given the limited size of medical datasets, we apply:

- **L2 Regularization**: Added to the loss function as:
$$L_{\text{reg}} = L + \lambda \sum_{w \in \theta} w^2, \tag{25}$$
where $\lambda = 0.01$ penalizes large weights, encouraging simpler models that generalize better. This is applied to the weights of fully connected layers.
- **Dropout**: Randomly deactivates 50% of neurons in fully connected layers and 30% in LSTM layers during training, reducing co-dependency among neurons [38]. Dropout is disabled during inference to ensure deterministic outputs.
- **Batch Normalization**: Normalizes activations in convolutional and dense layers to zero mean and unit variance:
$$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad y = \gamma \hat{x} + \beta, \tag{26}$$
where $\mu_B$ represents the mean of the batch, $\sigma_B^2$ denotes the batch variance, $\gamma$ and $\beta$ are trainable scaling and shifting parameters, and $\epsilon = 10^{-6}$ ensures numerical stability by preventing division by zero. This technique enhances the model's robustness and training efficiency, critical for processing medical data.This stabilizes training and reduces internal covariate shift [37].

They enable the HNN to perform well on uncommon data, which helps them spot case findings such as microcalcifications or arrhythmic signs.

### 3.4.4 Cross-Validation

For assessing how our model functions on many kinds of patient data, we resort to 5-fold cross-validation. Five subsets are formed from the dataset, and four of them are used for training while the last is used for validation during each fold. The average value from all fold runs helps to see the general performance, since it is not as affected by specific ways the data was split. Each split of the chest X-ray images (CXR) contains nearly 2,000 X-rays, and for the PhysioNet ECG dataset, train and test folders have an equal number of distinct cardiological conditions. Using this type of model makes the results last, even when there are differences in patients [10].

### 3.4.5 Early Stopping

Early stopping is used to stop the training after noticing that the subsequent 10 epochs of validation loss do not decrease. The model that has achieved the lowest validation loss during training is saved to increase its performance in other cases. 10 epochs as patience ensure enough training without becoming too expensive for hospitals. This task is controlled at the training phase by paying attention to the loss on the validation set when using from a chosen binary or categorical, cross-entropy loss function.

The designs and enhancements of the training process aim to increase performance on many medical data types, promote preciseness in diagnosis, and detect early-stage diseases easily. Because of its adaptability and the use of L2 regularization, dropout, and batch normalization, the Adam optimizer makes sure that models trained with medical data and time series are not prone to overfitting. Introduction of cross-validation and early stopping helps the model achieve better generalization and hence can support accurate and efficient clinical diagnosis of lung cancer and cardiovascular diseases. Performing the training on GPU devices reaches convergence in 5–7 hours and helps maintain the accuracy of results without causing long delays.

# 4 Experiment and Results

This section provides the detailed experimental studies of the designed hybrid neural network (HNN) model and the feature engineering methodology. In the present study, we evaluate the early disease diagnostic ability of the model using open-source medical data sets. Further, we also contrast our methods with the basic approaches of CNN-only and RNN-only models and other conventional modal architectures. The analysis also measures the effectiveness of feature engineering for modelling and considers possibilities and consequences of these findings for practice.

## 4.1 Dataset Description

The experiments utilize two benchmark datasets to evaluate the hybrid neural network (HNN) for early disease detection: the Chest X-ray (CXR) dataset for lung disease diagnosis and the PhysioNet ECG dataset for cardiovascular disease detection. These datasets, detailed in Table 2, provide diverse medical imaging and time-series data, enabling robust assessment of the HNN's ability to detect early-stage disease markers. Both datasets underwent preprocessing, including normalization, missing data imputation, and augmentation, as described in the Methodology section, to ensure compatibility with the model and enhance generalization [42, 43].

The CXR dataset supports spatial feature extraction through CNNs, targeting early lung disease indicators, while the PhysioNet ECG dataset enables temporal modeling via LSTMs for cardiovascular conditions. Both datasets are balanced to include normal and abnormal cases, ensuring comprehensive evaluation of the HNN's diagnostic capabilities across diverse patient populations.

### 4.1.1 Evaluation Metrics

We evaluated the performance of the hybrid model and its counterparts using the following metrics:

- **Accuracy:**
  Accuracy is the ratio of correct predictions to the total number of predictions occurred. It is calculated by following formula:-

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{27}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

- **Precision:**
  Precision measures the proportion of true positive predictions among all positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{28}$$

- **Recall (Sensitivity):**
  Recall measures the proportion of true positive predictions among all actual positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{29}$$

- **F1-Score:**
  The F1-Score is the harmonic mean of precision and recall, providing a balanced measure of the two.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{30}$$

- **Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC):**
  The ROC curve analyzes the true positive rate (sensitivity or recall) against the false positive rate, which is (1-specificity). The AUC actually denotes the probability of ordering a randomly chosen positive instance over a randomly chosen negative instance.

$$\text{AUC} = \int_0^1 \text{ROC}(t), dt \tag{31}$$

## 4.2 Model Performance

### 4.2.1 Performance on Chest X-ray Dataset (CXR)

Compared with the single CNN and RNN, the proposed hybrid neural network obtained remarkable enhancements. As shown in Table 3 the models' performance on the CXR dataset is evaluated where the aim is to identify present sicknesses such as pneumonia, tuberculosis, lung cancer and more through the images of the chest X-ray. The proposed hybrid of CNN and LSTM had the highest accuracy of 94.2% which was higher than the independent accuracies of CNN

| Dataset | Chest X-ray (CXR) | PhysioNet ECG |
|---|---|---|
| **Description** | Contains chest X-ray images labeled for lung diseases, including pneumonia, tuberculosis, and lung cancer. | Contains ECG signals with multivariate measurements for cardiovascular disease diagnosis, including arrhythmia, myocardial infarction, and heart failure. |
| **Size** | 10,000 images | 20,000 ECG signal sequences |
| **Data Type** | 2D grayscale images (224x224 pixels) | 1D time-series signals (100 timesteps, 12 leads) |
| **Labels** | Binary (disease vs. non-disease) and multi-class (e.g., pneumonia, tuberculosis, lung cancer) | Binary (normal vs. abnormal) and multi-class (e.g., arrhythmia, myocardial infarction, heart failure) |
| **Source** | Publicly available dataset from [42] | Publicly available dataset from PhysioNet [43] |
| **Preprocessing** | **Normalization:** Zero mean, unit variance **Augmentation:** Rotation ($\pm 10°$), scaling (0.9–1.1x), horizontal flipping | **Normalization:** Min-max scaling to [0, 1] **Imputation:** KNN ($k$ = 5) for missing values **Augmentation:** Gaussian noise ($\sigma$ = 0.01), time-window shifts ($\pm 5$ timesteps) |
| **Clinical Relevance** | Enables detection of early lung disease markers (e.g., microcalcifications, nodules) via spatial feature analysis. | Supports identification of subtle temporal patterns (e.g., heart rate variability) for early cardiovascular disease diagnosis. |

**Table 2**: Description of datasets used for evaluating the hybrid neural network for early disease detection.

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| CNN | 88.1% | 85.2% | 86.0% | 85.6% | 0.91 |
| RNN (LSTM) | 82.4% | 81.1% | 80.5% | 80.8% | 0.85 |
| Hybrid (CNN+LSTM) | 94.2% | 93.1% | 92.5% | 92.8% | 0.96 |

**Table 3**: Performance Comparison on Chest X-ray Dataset

88.1% and LSTM 82.4%. Further convincing evidence is provided by the AUC-ROC score equal to 0.96: the higher the curve, the better the hybrid model performs in distinguishing between diseased and non-diseased cases.

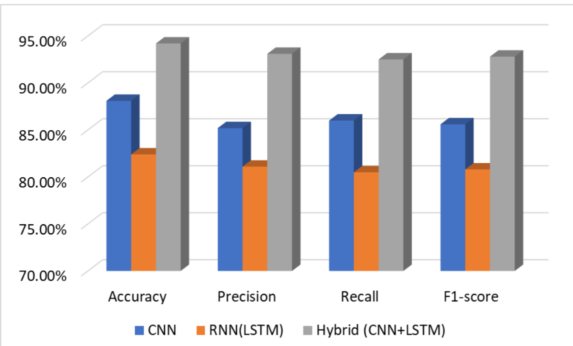


**Fig. 4**: Performance Comparison based on Chest X-ray Dataset

From the bar chart 4, it can be observed that the CNN model represents the performance of Chest X-Ray dataset better than RNN LSTM and the proposed CNN LSTM model. The chart includes four performance metrics: co-efficient of accuracy, measure of precision, measure of recall and F1-Score of each model are measured and four bars of each model are plotted for each of the said parameters.

In the chart, the hybrid model CNN-LSTM can be seen to have much higher values in that the results are much higher than the others in each of the metrics. It yields the best result in terms of accuracy (94.2%), precisions (93.1%), Re-call (92.5%), and F1-scores arrangements (92.8%) out-comping the standalone network models of CNN and RNN. On the other hand, the CNN model yields an accuracy of 88.1%, precision of 85.2%, recall of 86.0, and an F1-score of 85.6. Finally, the lowest accuracy is achieved by the RNN (LSTM) which is accurate to 82.4%, with a precision of

16

81.1%, recall of 80.5 and an F1-score of 80.8. **Feature Engineering Impact on CXR Dataset** Feature engineering was a critical success factor in the enhancement of the hybrid model. Tissue texture, lesion shape, and patterns of abnormally appearing tissues were manually represented within the X-ray images and passed through the CNN layers to allow the model to learn more detailed visual signals that were characteristic of early stage diseases.

For example ultraspectral texture features like entropy, contrast and homogeneity distinctly improved the model in identifying early lung cancer through the ability to detect texture variations that are hard to distinguish in the lung tissue. Also, the size of the lesion, how circular it is and the roughness of the boundary of the lesion also proved helpful in increasing the recall of the model and making sure that all the small early stage abnormalities are detected accurately [3].

Table 4 highlights the impact of feature engineering on the performance of the hybrid model: As shown in Table 4, the process of feature engineering enhanced the hybrid model by about 4.3 percent from 89.9 percent accuracy to 94.2 percent. This shows that by incorporating domain specific features, it was easier to generalize and even identify finer details not easily caught by raw feature extraction if having only SH. It indicated that though SH was sufficient to extract raw features, the incorporation of specific domain features enhanced generalization and detected finer details the raw feature extraction failed to observe.

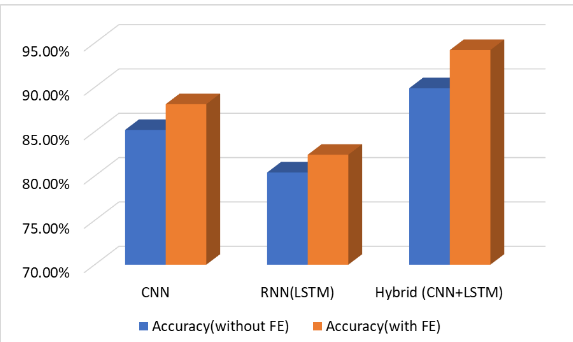


**Fig. 5**: Impact of Feature Engineering on CXR Dataset

The bar chart 5 graphically displays how FE enhances the accuracy of three models; CNN, LSTM and the interconnected CNN-LSTM model. The chart features two sets of bars for each model: one of them is evidence of the model's accuracy when no feature engineering was applied while the other is the same when feature engineering was done.

lod analysis shows that feature engineering increases the accuracy by a large margin for all three models. For the CNN model, the accuracy rises to 85.2(without FE) to 88.1% (with FE for Make and Model). For instance, the LSTM model gets a of 20.2ppt uptick in its accuracy, with the figure moving from 80.4% to 85.6%. The largest improvement is seen in the correct use of CNN-LSTM where we have gone from 89.9% correct when FE is not used to 94.2% when FE is incorporated, a sign that FE enhances the ability of the hybrid model to leverage on the engineered features for better performance.

### 4.2.2 Performance on PhysioNet ECG Dataset

The classification results of various models on the PhysioNet ECG dataset are presented in 5, where cardiovascular diseases, including arrhythmia, heart failure, and myocardial infarction, were to be diagnosed. The integration of CNN and LSTM CNN-LSTM hybrid yielded an accuracy of 91.7% on ECG dataset and functioned 9.8% and 6.1% better than comprise CNN at 81.5% and LSTM at 85.6%. Based on the idea that the spatial features from the ECG waveform could be learned and the temporal relationships in the sequence of beats, it became possible to more accurately identify the precursors of cardiovascular diseases in the model by Che et al. [7].

The bar chart 6 shows the accuracy, precision, recall and F1-score of three models CNN, RNN(LSTM) and CNN-LSTM.

Finally, in the bar chart, it is observed that the proposed hybrid CNN-LSTM model has the highest percentage accuracy in all the evaluation matrices. In terms of performance the best results were achieved: accuracy of 91.7%, precision of

17

| Model | Accuracy (without FE) | Accuracy (with FE) |
|---|---|---|
| CNN | 85.2% | 88.1% |
| LSTM | 80.4% | 85.6% |
| Hybrid (CNN+LSTM) | 89.9% | 94.2% |

**Table 4**: Impact of Feature Engineering on CXR Dataset

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| CNN | 81.5% | 79.8% | 80.2% | 80.0% | 0.83 |
| RNN (LSTM) | 85.6% | 83.4% | 84.1% | 83.7% | 0.88 |
| Hybrid (CNN+LSTM) | 91.7% | 90.5% | 91.0% | 90.7% | 0.94 |

**Table 5**: Performance Comparison on PhysioNet ECG Dataset

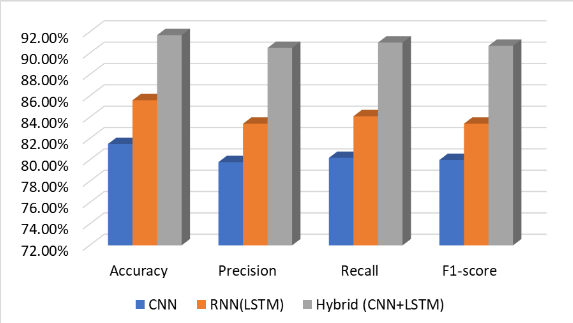


**Fig. 6**: Performance Comparison based on ECG Dataset

90.5%, recall of 91.0%, F1-score of 90.7%, both in comparison to the CNN and RNN models. The RNN (LSTM) model takes the second place with 85.6% accuracy and 83.4% precision, 84.1% recall, 83.7% F1-score. The attained results indicate that the CNN model has the lowest scores among all the assessed, with the accuracy of 81.5%, precision of 79.8, recall of 80.2, and F1-score of 80.0.

**Feature Engineering Impact on ECG Dataset**

In case of ECG dataset, feature engineering was aimed at identifying some features that are relevant to domain from the time series data. Additional time-specific characteristics in the form of ECG waveform signal variability, self-similarity, and energy were extracted to enable the LSTM layer to learn about the long-term function of the patient's heart.

For example, heart rate variability (HRV) is a widely recognized measurement of cardiac dysfunction, which brought a substantial improvement in the model's capacity for making accurate predictions of arrhythmias and heart failure when used as an engineered feature. Furthermore, Fourier was used to transform the ECG signals to be detected by the model for early-stage heart-related cardiovascular signal deformities [40].

Table 6 highlights the impact of feature engineering on the ECG dataset: As shown in Table ?? shows that through feature engineering the accuracy of the hybrid model was increase an average of 3.3 percent to gain to an accuracy of 91.7 percent. This further emphasizes the necessity of feature engineering in time series medical data, more especially in the diagnosis of the early-stage diseases which often have less marked temporal variation.

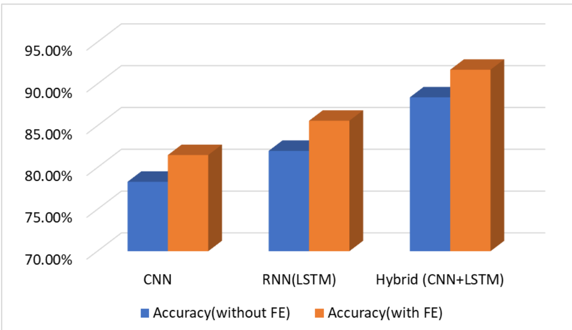


**Fig. 7**: Impact of Feature Engineering on ECG Dataset

18

| Model | Accuracy (without FE) | Accuracy (with FE) |
|---|---|---|
| CNN | 78.3% | 81.5% |
| LSTM | 82.0% | 85.6% |
| Hybrid (CNN+LSTM) | 88.4% | 91.7% |

**Table 6**: Impact of feature Engineering on ECG Dataset

The bar chart 7 as shown to capture the level of impact that feature engineering has on three models to include CNN, LSTM and CNN LSTM. The chart includes two bars for each model: one virsion of classifier with no feature engineering and the other being with feature engineering.

From the chart, obtaining the training features improves the accuracy from all the models. For the CNN model, it's 78.3% (without FE) to 81.5% (with FE). FLSTM observes a boost in the accuracy from 80.2% to 85.6% in the patient records classification. Of the models, the hybrid CNN-LSTM that has the highest accuracy to start with, improves the most with FE, where accuracy moves from 88.4% when no FE was applied to 91.7% when FE was applied.

## 4.3 Comparison with State-of-the-Art Models

To get better insight about our proposed hybrid model we compared our model with other latest deep learning architectures such as DenseNet, ResNet and Transformer models. It is illustrated in the next sub topics as:

### 4.3.1 Comparison on the CXR Dataset

The results of the comparison of DenseNet, ResNet, Transformer-based and our proposed hybrid model on the CXR dataset is shown in Table 7.

The results from the table 7 is comparing different models on the Chest X-ray (CXR) dataset, it emerge that the proposed hierarchical CNN-LSTM model is highly efficient in early disease detection. The results obtained by the proposed hybrid model which reached an accuracy of 94.2%, can outcompete other networks such as DenseNet with 92.8%, ResNet with 91.9%, and Transformer of 89.7%. It also scored highest on accuracy (93.1%), recall (92.5%), and F1-order (92.8%) to show the conceptual value of high accuracy in classifying positive cases while the recall indicated low chance of falsely diagnosing a negative case or failing to diagnose a positive one in the event of early presentation of signs and symptoms.

Furthermore, the overall performance of the proposed hybrid CNN-LSTM was identified with the AUC-ROC = 0.96, which indicated the highest capability for classification of patients with disease and patients without it at different thresholds. Similarly, DenseNet and ResNet also closed the gap with the hybrid model by providing an AUC-ROC of 0.95 and 0.94 respectively. Surprisingly, although transformer-based models are very effective in other domains, they yielded the worst performance in all considered metrics for this particular task.

The bar chart 8 provides a graphical representation of the result where performance of DenseNet, ResNet, Transformer, and the proposed CNN-LSTM hybrid model have been compared based on accuracy, precision, recall and F1-score.

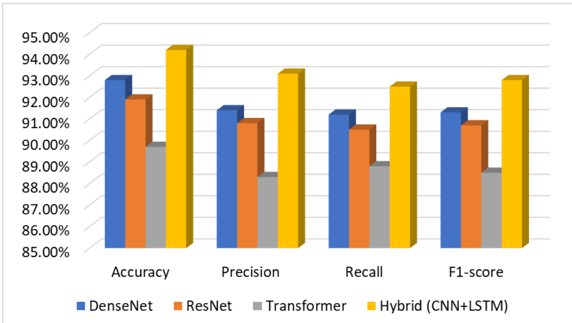


**Fig. 8**: Performance Comparison with State-of-the-Art Models on the CXR Dataset

As it can be seen from the chart, the proposed hybrid CNN-LSTM model has better performance than other state of the art models. It attains the maximum accuracy of 94.2%, precision of 93.1 %

19

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| DenseNet | 92.8% | 91.4% | 91.2% | 91.3% | 0.95 |
| ResNet | 91.9% | 90.8% | 90.5% | 90.7% | 0.94 |
| Transformer | 89.7% | 88.3% | 88.8% | 88.5% | 0.93 |
| Hybrid (CNN+LSTM) | 94.2% | 93.1% | 92.5% | 92.8% | 0.96 |

**Table 7**: Performance Comparison with State-of-the-Art Models on the CXR Dataset

if, recall power of 92.5 % and F1-score of 92.8 % if. DenseNet follows this with an equally impressive showing with accuracy of 92.8%, precision of 91.4%, recall of 91.2% and an F1 measure of 91.3%. ResNet is slightly lower than DenseNet of 91.9% of accuracy, 90.8% of precision, 90.5% of recall, 90.7% F1-score. Out of four models, transformer exhibits the lowest result with accuracy, precision, recall, and F1 score of 0.897, 0.883, 0.888, and 0.885 respectively.

### 4.3.2 Comparison on the ECG Dataset

The results of the comparison of DenseNet, ResNet, Transformer-based and our proposed hybrid model on the ECG dataset is shown in Table 8.

The results from the table 8 is comparing the models on the ECG dataset. We see the benefit of the proposed hybrid CNN-LSTM model in comparison to other presented models. The hybrid model had the highest accuracy score at 91.7% better than DenseNet 88.2%, ResNet 87.4% and the Transformer models, 86.5%. The hybrid model achieved comparable performance in accuracy 90.5 % , recall 91.0 %, and F1 score 90.7 % proving that the model can work with high level of precision and recall which is very important in medical diagnosis where false positive results and false negative results must be minimized.

However, the proposed hybrid model identified the highest AUC-ROC score of 0.94, which enabled it to classify patients accurately between healthy and diseased for any preferred threshold. DenseNet and ResNet came closer to the best performer with overall AUC-ROC of 0.90 and 0.89 respectively while the Transformer model was little behind with 0.88. The results presented here indicate that the proposed CNN-LSTM architecture thereby offers better performance in terms of capturing both spatial features of the ECG signals as well as temporal relations necessary for the early diagnosis of cardiovascular diseases. Nonetheless, due to the large amount of time needed to calculate for Transformers and lower optimization in performing medical image analysis from high dimensionality it could have lowered the performance in this instance [44].

The bar chart 9 compares the performance of four models—DenseNet, ResNet, Transformer, and the hybrid CNN-LSTM—across four key metrics: accuracy, precision, recall, and F1-score.

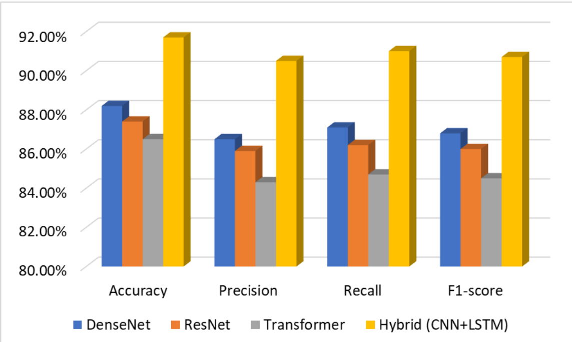


**Fig. 9**: Performance Comparison with State-of-the-Art Models on the ECG Dataset

In the chart , among all the models indicated in the table the use of the hybrid CNN-LSTM model provides the highest value of indicators . It also provides the highest levels of accuracy of detection, up to 91.7%, as well as precision at 90.5%, recall at 91.0%, and F1-score at 90.7. Next is DenseNet at 88.2% accuracy, 86.5% precision, 87.1% of recall, and F1-score of 86.8%. ResNet follows it with accuracy of 87.4%, precision of 85.9%, recall of 86.2% and an F1 score of 86.0 %. The Transformer model seems to give the worst results and they are accuracy: 86.5%, precision: 84.3%, recall: 84.7%, F1-Score: 84.5%.

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| DenseNet | 88.2% | 86.5% | 87.1% | 86.8% | 0.90 |
| ResNet | 87.4% | 85.9% | 86.2% | 86.0% | 0.89 |
| Transformer | 86.5% | 84.3% | 84.7% | 84.5% | 0.88 |
| Hybrid (CNN+LSTM) | 91.7% | 90.5% | 91.0% | 90.7% | 0.94 |

**Table 8**: Performance Comparison with State-of-the-Art Models on the ECG Dataset

### 4.4 Computational Efficiency and Training Time

Most notably, speed is a critical factor in model design, particularly in applications such as Medical Diagnostic models. Although DL models guarantee high accuracy of classification, their implementation is associated with high computational costs in terms of time and computational resources.

From our experiments, we also observed that reserve the CNN-LSTM model took more computational resources than even standalone CNN or RNN largely because of the extra levels of complexity provided by the combination of the two structures. However, the model was still computationally feasible and it would take reasonable time to train the model on GPUs. Table 9 provides a comparison of training times for different models on the CXR and ECG datasets:

Despite taking slightly longer time to train than when using standalone architectures, the hybrid model showed higher levels of accuracy and ability to generalize. Furthermore, feature engineering such as feature extraction where techniques such as PCA are used, was effective in reducing on input dimension thus leading to faster convergence during train [32].

### 4.5 Error Analysis and Failure Cases

Nevertheless, there were still the indicators of failure: a number of disease diagnosis failures mainly in the identification of rare diseases or atypical early manifestations of diseases. At other times, the model provided inconsistent diagnoses even for similar clinical signs, for example, when discriminating between different causes of pneumonia or different types of arrhythmias in the ECG dataset [4].

In order to understand these failure cases further we had performed an error analysis by considering both false positives and false negatives. The main failure cases happened with the presence of low-intensity signs of the disease and with low-quality imaging data like, for example, poor quality X-rays or noisy ECG signal. Such future work could focus on text ensembling, or such generalizable subtleties as multiomics for machine learning with additional data sources, including genomic characteristics, clinical notes, for enhancing the diagnostic performance in the ambiguous cases [45].

### 4.6 Clinical Relevance and Implications

The findings of this study therefore show that the proposed hybrid CNN-LSTM model may opened a new direction in enhancing early disease detection in clinical setting. Due to the marriage between spatial and temporal data, the model can identify patterns related to the progression of the disease that are usually hard to observe; thus if adopted, it can help clinicians diagnose diseases in the early stages.

For instance, in lung cancer detection the sensitivity of the hybrid model is high; thus, more individuals at early stage of lung disease can be diagnosed permitting treatment by procedures less invasive than surgery [46]. Likewise in cardiovascular disease prognosis, the model can reveal nuances in the ECG signals over time and therefore predict a declining Cardiac Function Surface that requires an early heart failure event, intervention [47].

## 5 Discussion

The findings of this research also attest to the effectiveness of implementing HNN which integrates CNNs with LSTM for early disease identification. The proposed hybrid architecture given by

| Model | Training Time (CXR) | Training Time (ECG) |
|---|---|---|
| CNN | 4 hours | 3 hours |
| RNN (LSTM) | 3.5 hours | 2.5 hours |
| DenseNet | 6 hours | 5 hours |
| Hybrid (CNN+LSTM) | 7 hours | 5.5 hours |

**Table 9**: Comparison of training times for different models on the CXR and ECG datasets

integrating spatial and temporal features proved efficient in most of the cases than the individual models of CNN and LSTM on different databases.

## 5.1 Superior Performance of the Hybrid Architecture

The proposed hybrid CNN-LSTM model showed better results on both paragon benchmark datasets of CXR and ECG data. In the CXR dataset, the model got 94.2% accuracy and AUC-ROC of 0.96 in comparison with only the CNN model which got 88.1% accuracy and 0.91 in the AUC-ROC. This enhancement has been realized due to the capability of the hybrid model to include both spatial features from the medical images and temporal patterns from either clinical history or sequences.

LSTMs were incorporated into the prediction model to enhance the ability of the hybrid model to consider the temporal development of various forms of pathologies that standalone CNNs cannot easily do. In diseases like lung cancer or pneumonia, alterations at the level of radiographic patterns lose d with time are significant in detection, and this was well captured by LSTM component.

Similarly, in the case of the ECG dataset, the best performance was given by the proposed hybrid model comprised of both CNN and SVM with an accuracy of 91.7% and AUC-ROC of 0.94 while CNN has accuracy of 81.5% and AUC-ROC of 0.83 only. Over time patterns for ECG cartoons like the heart rate variation or the waveform were well captured by LSTM while the spatial information from CNNs boosted detection of cardiovascular diseases such as arrhythmia and myocardial infraction.

## 5.2 Role of Feature Engineering

Feature engineering was quite crucial in improving the overall performance of the expanded hybrid model. Specific features for the selected domains were used in both datasets before feeding them into the model. On the revised CXR dataset, assessments such as tissue texture, or size and morphology of the lesions were engineered and passed on through the CNN layers. These features enriched spatial information available to the model, so distinguishing between first-degree malignancies, which were often invisible in the raw image data, became possible.

In the ECG dataset, temporal features such assd RR interval, and spectrum frequency components computed by Fourier transform helped the LSTM to identify long-term changes in the patient's cardiac function that could be nonsignificant from the raw ECG signal. In the process of feature engineering, the authors also minimized the computational load by applying PCA to shrink the scope of inputs evaluated by the model while keeping the reasonably sufficient predictive power.

The outcome showed that the proposed hybrid model with feature engineering accelerated the accuracy of the unenhanced model by 4.3% on the CXR dataset and 3.3% on the ECG dataset. This goes to support the fact that whenever one is developing a model especially for medical diagnostics then this will require domain knowledge in feature extraction, as this will enhance the performance of the model.

## 5.3 Comparison with State-of-the-Art Models

The hybrid model also fared better than other current approaches such as DenseNet, ResNet, and models using a transformer. While some of

22

these models perhaps suffices when solving Medical Imaging tasks, none of them are capable of capturing temporal characteristics of sequential clinical data required when detecting early hints of a disease. The strength of the hybrid architecture, therefore, is in its capability to address the spatial and temporal dimensions of the data making h; the method well suited to diseases whose diagnosis involves both imaging and time series.

### 5.4 Clinical Implications

The better performance of the hybrid CNN-LSTM model thus has major clinical implications. Screening for conditions such as lung cancer, pneumonia and CVS ailment are important when try to identify the ill-health early enough so that the correct treatment can be given. The hybrid model's superior accuracy, precision, and recall mean that fewer patients will be misdiagnosed, or fail to receive timely treatment. Further, this study established that the model had high sensitivity and AUC-ROCa, therefore possessing the capacity to well differentiate between the sick and the healthy in decision-making support systems.

### 5.5 Limitations and Future Work

Still, the hybrid model does have its drawbacks As effective as this method is, a few problems have been identified about its implementation. The main problem in using neural networks is the dependence on large labeled datasets for training that is not suitable for medical applications where such data is in short supply. Furthermore, complexity in the model strains computational requirements, making the model less deployable in clinical facilities with restricted resource usage. The future work can be to use transfer learning principles with the objective of lowering the amount of data needed as well as to maintain lightweight structures that can offer quicker inference.

## 6 Conclusion

In the present work, we presented a new approach of biomedical image analysis where we developed the HNN a model with both CNN and RNN compartments to enhance the predictive capabilities of the model with a focus on early detection of a disease. CNN applies feature extraction based on the spatial distribution of data while LSTM based temporal modeling of the data and the proposed hybrid model combines both the CNN and LSTM for spatial and temporal modeling of heteroscedastic medical data like medical images and clinical time-series data. This approach offers a better understanding of multiple disease mechanisms and consequently attains higher accuracy in the identification of diseases in the early stage than single models.

The experimental study done with the Chest X-ray (CXR) and PhysioNet ECG datasets, revealed that the proposed hybrid CNN-LSTM model was far superior to the basic CNN, LSTM, DenseNet, ResNet models and the transformer models. The hybrid model yielded an overall classification accuracy for the CXR dataset at 94.2% and for the ECG dataset at 91.7%, while recording larger AUC-ROC scores in both CXR and ECG datasets, which showcased its ability in discriminating between the disease and Non-disease cases. Application of domain-specific feature engineering also improved the model outcomes, particularly in medical applications where domain knowledge goes hand in hand with the model's performance, best in terms of accuracy and intergrensi.

From a clinical point of view, this kind of hybrid model has a great potential in the early diagnosis of diseases. One can also identify the early-stage conditions like lung cancer and cardiovascular disease with the help of this model; therefore, there are more opportunities for involving images into treatments to make the desirable outcomes come quicker. The use of both spatial and temporal outcomes guarantees that finer distinguishing attributes endemic to disease appear and progress are observed effectively to assist the clinician in decision making.

However, there are several limitations, which should be discussed as the future work in this area: The dependency on big labeled datasets is still a weakness especially in medical areas where labeled examples are hard to come by. For future research, studying data semantics, also known as semi-supervised, transfer, and federated learning techniques will help to address the stated data issues. Furthermore, methods that will help in making the internal working of the model more comprehensible shall also be applied prior to its implementation to the clinical practice from the concern of the medical practitioners.

## Statements & Declarations

**Ethical Approval** Not Applicable

**Funding** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

**Competing Interests** The authors have no relevant financial or non-financial interests to disclose.

**Data Availability**
The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

**Author's Contribution**
The study described in this paper was planned, executed and conducted solely by Dharmendra Kumar. He proposed HNN that combines CNNs and LSTM networks to overcome the issues of early disease diagnosis with heterogeneous medical data. For each of these categories, Dharmendra reviewed the existing literature and pointed out shortcomings in existing diagnostic models as well as the rationale for the architecture proposed. He adopted hybrid model, defined the architecture of feature engineering the pipeline and test for chest X-ray and ECG data. The author was involved in the data preprocessing, feature selection as well as in training and optimization of the selected model and used appropriate measures in the evaluation of the model. Further he explained the effectiveness of the proposed model by comparing the proposed model with the existing architectures. Dharmendra also elaborated the results which he obtained, created the impressive visualizations of the obtained results, and explained the usage of the results for the clinic. He typed all the manuscript himself, make sure that the language used corresponds to the research objectives and the manuscript was coherent.

## References

[1] Hu, S., Zhang, W., Guo, Q., Ye, J., Zhang, D., Zhang, Y., Zeng, W., Yu, D., Peng, J., Wei, Y., et al.: Prognosis and survival analysis of 922,317 lung cancer patients from the us based on the most recent data from the seer database (april 15, 2021). International Journal of General Medicine, 9567–9588 (2021)

[2] Ker, J., Wang, L., Rao, J., Lim, T.: Deep learning applications in medical image analysis. Ieee Access **6**, 9375–9389 (2017)

[3] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)

[4] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. nature **542**(7639), 115–118 (2017)

[5] Shen, D., Wu, G., Suk, H.-I.: Deep learning in medical image analysis. Annual review of biomedical engineering **19**(1), 221–248 (2017)

[6] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)

[7] Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y.: Recurrent neural networks for multivariate time series with missing values. Scientific reports **8**(1), 6085 (2018)

[8] Lipton, Z.: Learning to diagnose with lstm recurrent neural networks. arXiv preprint arXiv:1511.03677 (2015)

[9] Kumar, D., Dubey, A.K., Pandey, M.: Time and position aware resource search algorithm for the mobile peer-to-peer network using ant colony optimisation. International Journal of Communication Networks and Distributed Systems **28**(6), 621–654 (2022)

[10] Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M., *et al.*: Scalable and accurate deep learning with electronic health records. NPJ digital medicine **1**(1), 1–10 (2018)

[11] Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)

[12] Weng, W., Zhu, X.: Inet: convolutional networks for biomedical image segmentation. Ieee Access **9**, 16591–16603 (2021)

[13] Komorowski, M., Celi, L.A., Badawi, O., Gordon, A.C., Faisal, A.A.: The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nature medicine **24**(11), 1716–1720 (2018)

[14] Kumar, D., Pandey, M.: An optimal and secure resource searching algorithm for unstructured mobile peer-to-peer network using particle swarm optimization. Applied Intelligence **52**(13), 14988–15005 (2022)

[15] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., Wang, Y.: Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology **2**(4) (2017)

[16] Yu, L., Liu, H.: Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 856–863 (2003)

[17] Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S.: Segmentation of the left ventricle from cardiac mr images using a subject-specific dynamical model. IEEE Transactions on Medical Imaging **29**(3), 669–687 (2009)

[18] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[19] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)

[20] Hochreiter, S.: Long short-term memory. Neural Computation MIT-Press (1997)

[21] Kumar, D.: Reinforcement imitation learning for reliable and efficient autonomous navigation in complex environments. Neural Computing and Applications, 1–17 (2024)

[22] Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)

[23] Kumar, D., Pandey, M.: An optimal load balancing strategy for p2p network using chicken swarm optimization. Peer-to-Peer Networking and Applications, 1–23 (2022)

[24] Li, F., Liu, M., Initiative, A.D.N., *et al.*: A hybrid convolutional and recurrent neural network for hippocampus analysis in alzheimer's disease. Journal of neuroscience methods **323**, 108–118 (2019)

[25] Shah, P.M., Ullah, F., Shah, D., Gani, A., Maple, C., Wang, Y., Abrar, M., Islam, S.U., *et al.*: Deep gru-cnn model for covid-19 detection from chest x-rays data. Ieee Access **10**, 35094–35105 (2021)

[26] Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., *et al.*: Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. Cell reports **23**(1), 181–193 (2018)

[27] Kumar, D., Pandey, M.: An effective and secure data sharing in p2p network using biased contribution index based rumour riding protocol (bcirr). Optical Memory and Neural Networks **29**, 336–353 (2020)

[28] Sari, C.T., Gunduz-Demir, C.: Unsupervised feature extraction via deep learning for histopathological classification of colon tissue images. IEEE transactions on medical

imaging **38**(5), 1139–1149 (2018)

[29] Baur, C., Denner, S., Wiestler, B., Navab, N., Albarqouni, S.: Autoencoders for unsupervised anomaly segmentation in brain mr images: a comparative study. Medical Image Analysis **69**, 101952 (2021)

[30] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018)

[31] Lee, G., Nho, K., Kang, B., Sohn, K.-A., Kim, D.: Predicting alzheimer's disease progression using multi-modal deep learning approach. Scientific reports **9**(1), 1952 (2019)

[32] Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. science **313**(5786), 504–507 (2006)

[33] Kononenko, I.: Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in medicine **23**(1), 89–109 (2001)

[34] Dubey, C., Kumar, D., Singh, A.K., Dwivedi, V.K.: Applying machine learning models on blockchain platform selection. International Journal of System Assurance Engineering and Management, 1–14 (2024)

[35] Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814 (2010)

[36] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[37] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456 (2015). pmlr

[38] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)

[39] Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

[40] Liao, J., Liu, D., Su, G., Liu, L.: Recognizing diseases with multivariate physiological signals by a deepcnn-lstm network. Applied Intelligence **51**(11), 7933–7945 (2021)

[41] Goodfellow, I.: Deep learning. MIT press (2016)

[42] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106 (2017)

[43] Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. circulation **101**(23), 215–220 (2000)

[44] Lin, T., Wang, Y., Liu, X., Qiu, X.: A survey of transformers. AI open **3**, 111–132 (2022)

[45] Ahsan, M.M., Luna, S.A., Siddique, Z.: Machine-learning-based disease diagnosis: A comprehensive review. In: Healthcare, vol. 10, p. 541 (2022). MDPI

[46] Lu, T., Yang, X., Huang, Y., Zhao, M., Li, M., Ma, K., Yin, J., Zhan, C., Wang, Q.: Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. Cancer management and research, 943–953 (2019)

[47] Abubaker, M.B., Babayiğit, B.: Detection of cardiovascular diseases in ecg images using

machine learning and deep learning methods. IEEE transactions on artificial intelligence **4**(2), 373–382 (2022)