Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.57, p. 993-1008. January/February, 2026.

DOI: https://doi.org/10.5935/jetia. v12i57.3154

RESEARCH ARTICLE OPEN ACCESS

# ENHANCING DISEASE DIAGNOSIS IN HEALTHCARE USING LSTM NETWORKS WITH OPTIMIZED FEATURE SELECTION: A COMPARATIVE RESEARCH ON HEART DISEASE, BREAST CANCER, AND LIVER DISEASE DATASETS

**Tina Sachdeva*[1], Priyanka Sharma[2], and Dr. Mehtab Alam[3]**

[1]Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Vasundhara Enclave, Delhi, India.
[2, 3]Acharya Narendra Dev College, University of Delhi, Govind Puri, Kalkaji, India

[1]http://orcid.org/0009-0003-8798-8719, [2]http://orcid.org/0009-0009-3341-6823, [3]http://orcid.org/0000-0001-7554-2160

Email: *tina.sachdeva@rajguru.du.ac.in, priyankasharma@andc.du.ac.in, mehtabalam@andc.du.ac.in

**ARTICLE INFO**

***Article History***
Received: December 17, 2025
Revised: January 10, 2026
Accepted: January 15, 2026
Published: February 28, 2026

***Keywords:***
LSTM model,
feature selection,
medical diagnosis,
heart disease,
breast cancer,
liver disorder

## ABSTRACT

Clinical diagnosis relies heavily on accurate and timely medical assessment. However, clinical data sets are typically large, incomplete, and noisy, therefore limiting the reliability of traditional models used to diagnose clinical conditions. To address these limitations, this study developed an integrated model combining an optimised feature selection strategy with a Long Short-Term Memory (LSTM) model to provide improved accuracy across three clinical condition benchmarks, namely heart disease, breast cancer, and liver disorder. The first step in this process was feature selection, which involved selecting only those features that were deemed clinically relevant. By doing this, the subsequent LSTM model was able to recognise temporal patterns within the patient's clinical data set more easily than it would have been able to do using the original clinical data set. Results from this work demonstrates that the proposed methodology resulted in higher accuracy compared to baseline classifiers and also to non-optimised LSTM models across all three data sets. Specifically, the proposed method had an accuracy of 91.5% for heart disease, 97.6% for breast cancer, with a Receiver Operating Characteristic Area Under the Curve (ROC-AUC) of 0.99, and 83.9% accuracy for liver disorder. Additionally, there was a clear improvement in terms of recall and F1-score. Overall, the results from this study demonstrate that the integration of a method for dimensionality reduction with a sequential learning approach produces a reliable and generalizable clinical diagnostic model that can support clinical decision-making in various health care settings.

## I. INTRODUCTION

The increasing reliance on Machine Learning (ML) and Artificial Intelligence (AI) has made them essential components of contemporary health care, especially regarding automated and enhanced early disease detection [1]. Timeliness of diagnosis is also a very critical factor: an untreated cardiovascular condition may result in sudden death, delayed breast cancer detection leads to a lower survival rate, and untreated liver abnormalities will almost always result in irreparable damage. Manual analysis of laboratory and imaging data is the traditional basis for the majority of the diagnoses in clinical practice. While these approaches are successful, they suffer from human variability, subjectivity, and scalability issues in dealing with large numbers of patient data sets. Cardiovascular disease, breast cancer, and chronic liver disease are among the top reasons for morbidity and mortality globally. Cardiovascular disease is referred to as a "silent killer" because patients typically present with symptoms only after extensive physical damage [2]. The relationship between early detection and survival rates is well established for breast cancer, the most common form of cancer among women [3]. In the context of advanced liver disease, particularly in patients with underlying hepatitis or cirrhosis, complex portal vein thrombosis may arise and require specialised surgical reconstruction.[4].

Analysing these conditions as part of a unifying analytical framework helps to identify whether a single intelligent model can be generalised to different pathological areas. One of the persistent challenges in clinical analytics is the presence of poor or redundant variable information. Examples of features that are highly discriminatory in heart disease include cholesterol levels, blood pressure, and electrocardiogram (ECG) signals, whereas many peripheral features may have minimal effect on predicting outcomes. Maintaining irrelevant data contributes to the problem of the curse of dimensionality, causing overfitting and degrading the model's ability to generalise [5].

Techniques for feature selection can help alleviate the effects of redundancy by identifying the most influential predictor variables, reducing computational costs while providing clinicians with greater clarity of the relevance of retained variables. There is another aspect that is often neglected in clinical analytics, and that is the nature of the temporal evolution of the disease. Most diseases occur gradually, and isolated static measurements may hide underlying trends. For example, rising liver enzyme levels over time may signal early liver disease even though individual readings remain within normal limits. Likewise, subtle changes in ECG patterns observed over hospital visits may indicate impending cardiovascular events. To analyse the relationships in the sequences of clinical data, this research uses LSTM networks [6], a type of Recurrent Neural Network that retains long-range contextual information about sequences via special types of gates. Figure 1 depicts the overall structure of an LSTM.

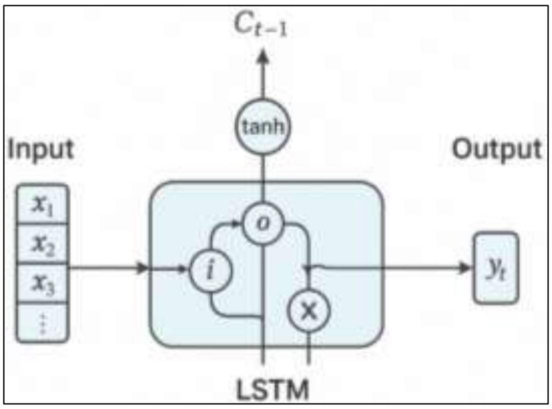

Figure 1: LSTM Network.
Source: Authors, (2026).

The work presented in this paper distinguishes itself from existing studies by introducing a tailored feature-selection approach followed by an LSTM model, and by evaluating their combined impact to enhance prediction performance across three publicly available datasets:

- **Heart Disease Dataset** (UCI Repository) [7]
- **Breast Cancer Dataset** (Wisconsin Diagnostic Dataset) [8]
- **Liver Disorder Dataset** (Indian Liver Patient Dataset) [9]

Maintaining consistency in the methodology used to evaluate each dataset allows this study to assess how optimisations to the feature space interact with temporal sequence modeling to improve clinical predictions. Maintaining a consistent evaluation methodology across all datasets enables this study to systematically examine how feature space optimisation interacts with temporal sequence modeling, ultimately contributing to improved clinical prediction performance.

## I.1 RESEARCH QUESTIONS

Additionally, the following key questions guide this research:

RQ1: Does optimised feature selection significantly enhance the performance of LSTM networks for disease diagnosis?

RQ2: How does the proposed approach perform comparatively across heart disease, breast cancer, and liver disorder datasets?

RQ3: Which evaluation metric (accuracy, precision, recall, F1-score, ROC-AUC) shows the greatest improvement after feature optimisation?

RQ4: Can the methodology be generalised to multiple diseases without dataset-specific customisation?

## I.2 CONTRIBUTIONS OF THIS RESEARCH

This research makes the following contributions:

- Proposes an **integrated framework** combining optimised feature selection with LSTM networks for medical diagnosis.
- Provides a **comparative evaluation** across three widely used benchmark datasets (heart disease, breast cancer, liver disease), highlighting generalizability.
- Demonstrates the role of **feature selection** in reducing computational complexity while improving classification accuracy.
- Offers **insights for healthcare practitioners** and data scientists on applying advanced deep learning methods in multi-disease contexts.

Table 1: Global Burden of the Targeted Diseases.

| Disease | Global Prevalence (approx.) | Mortality Impact | Key Challenge in Diagnosis |
|---|---|---|---|
| Heart Disease [7] | 523 million cases (2021) | ~18 million deaths annually | Often asymptomatic until late stages |
| Breast Cancer [8] | 2.3 million cases (2020) | ~685,000 deaths annually | Requires early detection for survival |
| Liver Disease [9] | 1.5 billion cases (2020, chronic liver conditions) | ~2 million deaths annually | Complex diagnosis due to multiple causes |

Source: Authors, (2026).

# II. LITERATURE REVIEW

## II.1 OVERVIEW AND SCOPE

This literature review focuses on two intersecting streams relevant to the proposed : (1) **deep sequence models** (particularly LSTM and hybrid LSTM architectures) applied to medical diagnosis (heart disease, breast cancer, liver disease and allied tasks), and (2) **feature selection and optimization** methods (filter, wrapper, hybrid, and metaheuristic optimizers such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) used to improve medical classification performance. Wherever available, preference is given to SCIE journals and high-quality open repositories (JAMA Network Open, Scientific Reports, BMC series, Expert Systems with Applications, Information Sciences, Sensors, etc.).

## II.2 KEY THEMES FROM RECENT LITERATURE

### II.2.1 LSTM and Hybrid LSTM Models for Medical Diagnosis

A growing number of studies deploy LSTM or hybrid Convolutional Neural Network (CNN–LSTM)/Conv1D–LSTM architectures for detection/classification tasks:

- Hybrid image-processing + sequence architectures (CNN + LSTM) show strong performance on breast cancer imaging and histopathology datasets [10]. Recent Scientific Reports and BMC Medical Imaging papers demonstrate near-state-of-the-art accuracy using CNN-LSTM hybrids.
- For cardiovascular or heart disease tasks, hybrid CNN–LSTM have been proposed to capture temporal patterns or sequential clinical features, with clear performance gains over shallow models [11].
- For liver disease and cirrhosis prognosis, deep sequence models have been used to model longitudinal clinical records and outperformed simple linear models in some cohort studies. JAMA Network Open and PLoS One showcase LSTM/RNN success in longitudinal hepatology prediction tasks [12].

LSTM (and LSTM hybrids) are widely validated across modalities — imaging (when combined with CNN front-ends), waveform data (ECG), and longitudinal clinical records — and routinely outperform classical linear/ensemble baselines when sequential information is present.

### II.2.2 Feature Selection (Classical and Optimised) in Medical Diagnosis

Feature selection remains a central preprocessing step in medical ML systems:

- Metaheuristic methods such as PSO and GA remain popular for wrapper-style feature selection in clinical tasks; Sensors (2021) and several recent ESWA / Information Sciences papers detail PSO/GA variants tailored to feature selection with improvements in accuracy and dimensionality reduction [13].
- Hybrid and multi-objective approaches that balance predictive performance and subset size (e.g., multi-objective GA, hybrid filter-wrapper pipelines) show improved robustness across datasets. Recent Applied Soft Computing / ScienceDirect works present dynamic and hybrid feature selection schemes for medical diagnosis [14].
- Newer filter+grouped/structured feature selectors and niche GA variants address high-dimensional gene expression/imaging features for cancer diagnostics, indicating that domain-aware selection (grouping correlated features) improves interpretability and stability [15].

Metaheuristic and hybrid selection methods are effective in producing compact, high-quality feature sets for medical classification. However, many works evaluate feature selection with classical classifiers (Support Vector Machine (SVM), Random Forest (RF)) rather than deep LSTM architectures.

### II.2.3 Combined Pipelines: Feature Selection + Deep Learning

Several recent papers examine integrating feature selection with deep models, but mostly in limited scopes:

- Some works fuse transfer-learning/CNN feature extractors with subsequent LSTM modules and perform either embedded or post-hoc selection to reduce sequence dimensionality for LSTM input (breast cancer, multimodal ultrasound) [16].
- A smaller number of studies explicitly optimise feature subsets for LSTM inputs using wrapper/metaheuristic search—but they focus on a single disease modality (e.g., heart or breast cancer) [17].

While there is precedent for combining feature selection and LSTM, comparative cross-disease evaluations using a consistent Feature Selection (FS) optimiser + LSTM pipeline are rare in high-quality SCIE literature.

### II.2.4 Evaluation Rigour and Clinical Considerations

- Several reviews and methodological papers emphasise the need for robust evaluation: cross-validation, statistical significance testing, attention to class imbalance, and bias analysis (e.g., sex/gender stratification in Indian Liver Patient Dataset (ILPD)) [18].
- The lack of interpretability in high-performing deep learning models is increasingly recognised as a barrier to clinical trust. Recent SCIE work, including attention-based architectures proposed, seeks to address this limitation by providing more transparent feature representations [19].

Table 2 summarises the selected recent works relevant to LSTM, hybrid LSTM architectures, and optimised feature selection in medical diagnosis.

Table 2: Summary of the literature review.

| Paper | Dataset / Modality | Method / Feature selection | Key result/relevance |
|---|---|---|---|
| [19] | Heart disease / clinical + big data | Attention-Gated Recurrent Unit (GRU) hybrid; data visualisation pipeline | Hybrid RNN with attention improves disease detection; shows explainability potential. |
| [10] | Breast imaging (mammograms) | CNN + LSTM for feature extraction + sequence modeling | Very high classification accuracy on benchmark imaging datasets; supports CNN-LSTM for breast cancer detection. |
| [20] | Histopathology images (breast) | Hybrid CNN + bi-LSTM | Achieved high binary accuracy (~99%) for benign/malignant classification; emphasises a hybrid approach. |
| [21] | Breast cancer (tabular & imaging) | New deep architecture (granular + shortcut connections) | Strong performance; reinforces deep model design diversity. |
| [16] | Multimodal ultrasound (breast) | Transfer learning fused with LSTM; dynamic feature fusion | Robust diagnosis with incomplete multimodal data; relevant for fusion with FS. |
| [11] | Cardiovascular (ECG / clinical) | CNN-LSTM hybrid; XAI components | Improved detection and interpretability for cardiovascular disease (CVD). |
| [22] | Review (multiple datasets) | Survey of deep learning in breast cancer | Synthesises imaging datasets, architectures, and gaps in interpretability. |
| [12] | Longitudinal liver records | RNN / LSTM modeling for risk prediction | RNN outperformed linear models for cirrhosis/HCC risk prediction. |
| [23] | Liver cirrhosis | Deep Learning (DL) and ML comparative analysis | ML/DL models outperform standard risk scores in some horizons. |
| [24] | Indian/clinical liver datasets | Ensemble ML + optimised FS | High accuracy (eXtreme Gradient Boosting (XGBoost)/Gradient Boosting (GB)) after FS; encourages DL extensions. |
| [13] | Methodological (general) | Enhanced PSO for feature selection | Demonstrated PSO variants improve selection stability and downstream classifier performance. |
| [25] | Cancer gene expression/imaging | GA recursive feature elimination | GA-based Recursive Feature Elimination **e**ffective for high-dimensional cancer data. |
| [26] | Mammography images | LSTM (sequence) + U-Net segmentation | The dual model improved tumour localisation and classification. |
| [27] | Medical classification (general) | Dynamic feature selection framework | Reported gains in classification accuracy vs static selection. |
| [28] | Medical datasets (varied) | Hybrid FS combining filter/wrapper | Robustness across disease datasets. |
| [29] | Clinical tabular datasets | Ensemble of learners + FS | Showed ensemble + FS improved diagnostic accuracy vs single models. |
| [30] | Medical diagnosis (general) | Novel optimiser for dynamic FS | Reported improved feature subset quality and classifier outcomes. |
| [21] | Breast cancer imaging/tabular | Deep LSTM/DNN variants | Reinforces the need for architecture experimentation. |
| [31] | Coronary heart disease (ECG & clinical) | Integration of ML models; preprocessing focus | Preprocessing-aware deep learning pipeline. |

Source: Authors, (2026).

## II.3 CROSS-PAPER SYNTHESIS AND METHODOLOGICAL TRENDS

- Hybrid pipelines outperform single-stage models: across imaging and sequential clinical data, combining feature extractors (CNN or Conv1D) with LSTM or attention-based RNNs yields consistent improvements [20]
- Feature selection increases robustness: PSO / GA / hybrid filter-wrapper approaches often reduce dimensionality and improve classical classifier performance; however, explicit wrapper optimisation targeted to deep LSTM inputs is less commonly reported in SCIE outlets [13].
- Evaluation practices vary: While many healthcare ML studies rely primarily on aggregate metrics such as accuracy and AUC, fewer explicitly examine demographic bias. This gap is highlighted in liver disease prediction by demonstrating that sex-stratified evaluation reveals performance disparities that are obscured by overall metrics, underscoring the need for bias-aware evaluation practices [32].
- Explainability is underdeveloped for LSTM pipelines: while some hybrid studies include attention maps or SHAP analyses, a persistent gap in producing clinically intelligible explanations for sequence-based predictions that map to known medical concepts [19].

## II.4 IDENTIFIED RESEARCH GAPS

From the surveyed literature, we identify the following gaps that the present research will address:

- Lack of systematic cross-disease comparative studies that apply the *same optimised feature selection + LSTM pipeline* to multiple, diverse clinical datasets (heart disease, breast cancer, liver disease, etc) under a unified evaluation framework. Many works focus on a single disease or modality.
- Limited wrapper optimisation targeted to LSTM inputs in SCIE outlets-While PSO/GA methods are well-studied for classical classifiers, few high-quality studies optimise feature subsets specifically for LSTM performance across multiple disease types. This is important because LSTM's sensitivity to input dimensionality and sequence structure differs from classical classifiers.
- Insufficient statistical rigour and bias analysis-Several recent papers omit formal statistical testing (e.g., paired t-test or Wilcoxon) across folds and do not analyse demographic biases (e.g., sex, age) that are known issues in public datasets such as ILPD.
- Explainability for sequence models is immature. There is a need to couple optimised FS + LSTM with interpretable attribution techniques that translate model decisions to clinical features meaningful to practitioners.
- Generalizability and reproducibility- Many high-performance pipelines are tested on single internal datasets or lack publicly available code and seeds, making comparisons and replication difficult. Cross-dataset evaluation with standard preprocessing and public datasets would improve reproducibility.

## II.5 HOW THIS RESEARCH WILL ADDRESS THE GAPS

To address the identified gaps, this research will:

- Design a consistent pipeline that pairs an optimised wrapper feature selection method (metaheuristic + stability constraints) with an LSTM classifier and apply it to three benchmark public datasets (Heart Disease, Wisconsin Breast Cancer, Indian Liver Patient Dataset).
- Evaluate using robust practices: stratified k-fold cross-validation, repeated experiments with fixed seeds, paired statistical significance testing (Wilcoxon / paired t-test where appropriate), and bias subgroup analyses (sex/age).
- Integrate explainability: post-hoc attribution (e.g., SHAP adapted for sequences, attention visualisation) to map selected features to clinical meaning.

# III. METHODOLOGY

The methodology of this research integrates optimised feature selection with Long Short-Term Memory (LSTM) networks for disease classification across three benchmark medical datasets: Heart Disease, Breast Cancer, and Liver Disorder. The working of the proposed model is designed to optimise the feature space and give better performance with raw data. We have also trained the LSTM model and then evaluated the performance of the model using several statistical metrics. Next, we have discussed all the stages in detail.

## III.1 DATA SOURCES

In this research, we have used three widely known benchmark datasets. These datasets ensure comparability and reproducibility of results:

### III.1.1 Heart Disease Dataset (Cleveland Subset, UCI Repository)

- Features: 13 attributes such as age, sex, chest pain type, cholesterol, maximum heart rate, etc.
- Samples: 303 patients.
- Target variable: Presence (1) or absence (0) of heart disease.

### III.1.2 Breast Cancer Dataset (Wisconsin Diagnostic Dataset, UCI Repository)

- Features: 30 attributes such as radius, texture, smoothness, symmetry, etc. extracted from digitised breast mass images.
- Samples: 569 patients.
- Target variable: Malignant (M) or Benign (B).

### III.1.3 Liver Disorder Dataset (Indian Liver Patient Dataset, UCI Repository)

- Features: 10 clinical attributes, including total bilirubin, alkaline phosphatase, Serum Glutamic-Oxaloacetic Transaminase (SGOT), albumin, etc.
- Samples: 583 patients.
- Target variable: Liver disorder present (1) or absent (0).

Table 3 depicts the characteristics of the three datasets used.

Table 3: Dataset Characteristics.

| Dataset | No. of Samples | No. of Features | Target Classes | Source |
|---|---|---|---|---|
| Heart Disease | 303 | 13 | Disease / No Disease | UCI Repository |
| Breast Cancer | 569 | 30 | Malignant / Benign | UCI Repository |
| Liver Disorder | 583 | 10 | Liver Disease / No Disease | UCI Repository |

Source: Authors, (2026).

## III.2 DATA PREPROCESSING

Medicine's raw datasets are noisy or incomplete and thus require preprocessing. To obtain quality input for our LSTM, we have performed the following:

- Missing Value Handling: Numerical attributes with missing values were imputed using median values to minimise distortion.
- Normalising: scales features to a specific range, typically [0,1]. This prevents attributes that naturally have large amplitudes (cholesterol levels) from overpowering smaller ones in prediction (age).
- Categorical Encoding: for non-numeric features, we used one-hot encoding (for example, “Chest Pain Type” with choices {typical angina, atypical angina, non-anginal pain, asymptomatic} is represented as four binaries).
- Data Splitting: each dataset was split into training (80 percent) and testing (20 percent) subsets. This gives a more honest estimate of performance.

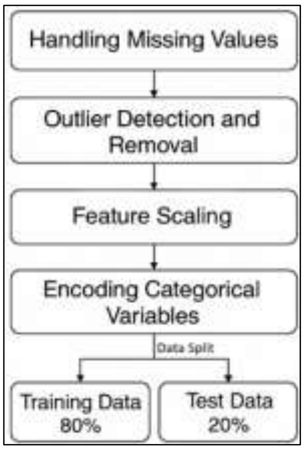

Figure 2: Data Preprocessing Pipeline.
Source: Authors, (2026).

## III.3 FEATURE SELECTION USING OPTIMISATION

High-dimensionality in datasets can produce excess features, irrelevant to the needs and detrimental to performance. We use metaheuristic optimisation algorithms for feature selection.

- Technique Chosen: Particle Swarm Optimisation (PSO), due to its applicability in analysing healthcare data
- How It Works: It takes its inspiration from the collective movement seen in flocks of birds. Each particle represents a certain subset of features. The algorithm iteratively updates the position of each particle, considering the best solution found by the particle itself and the best solution found by any particle in the neighbourhood.
- Objective Function: Fitness of each subset of features is measured by the classification accuracy of a lightweight machine learning model (logistic regression). We selected the subset that achieved the best fitness score for the "real" LSTM.

## III.4 LONG SHORT-TERM MEMORY (LSTM) MODEL

LSTM networks are an extension of recurrent neural networks (RNNs) designed to overcome the problem of vanishing gradients. The model consists of memory cells and three types of gates: input, forget, and output.

- Input Gate: Decides which values from the current input should be stored in memory.
- Forget Gate: Determines which past information should be discarded.

- Output Gate: Selects which information from memory will be passed forward.

Architecture used for this Research:

- Input layer: Dimension equal to optimised feature set size.
- Two hidden LSTM layers: Each with 64 and 32 units, respectively.
- Dropout: 0.3 applied to reduce overfitting.
- Fully connected dense layer: With sigmoid activation for binary classification.
- Optimiser: Adam optimiser with learning rate 0.001.
- Loss function: Binary cross-entropy.

## III.5 MODEL TRAINING AND EVALUATION

- Training: Each dataset was trained for 100 epochs with a batch size of 32. Early stopping was applied to prevent overfitting.
- Cross-Validation: 10-fold cross-validation was employed to ensure statistical robustness.

### III.5.1 Evaluation Metrics

- Accuracy: Percentage of correct predictions.
- Precision: Ratio of correctly predicted positives to total predicted positives.
- Recall (Sensitivity): Ability to correctly identify true positives.
- F1-Score: Harmonic mean of precision and recall.
- ROC-AUC: Area under the Receiver Operating Characteristic curve.

Table 4 lists the details of the above-mentioned evaluation metrics.

Table 4: Evaluation Metrics Formulation.

| Metric | Formula | Interpretation |
|---|---|---|
| Accuracy | $\frac{TP+TN}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$ | Overall correctness of predictions |
| Precision | $\frac{TP}{(\mathrm{TP}+\mathrm{FP})}$ | Proportion of relevant positive cases detected |
| Recall | $\frac{TP}{(\mathrm{TP}+\mathrm{FN})}$ | Ability to detect all positive cases |
| F1-Score | $2\times\frac{Precision\times Recall}{Precision+Recall}$ | Balance between precision and recall |
| ROC-AUC | Integral under the ROC curve | Discrimination capability between classes |

Source: Authors, (2026).

## III.6 EXPERIMENTAL WORKFLOW

The complete experimental workflow is summarised below:

- Data acquisition from UCI repositories.
- Preprocessing: Missing value handling, normalisation, encoding, and data splitting.
- Feature selection using PSO.
- Model training with LSTM architecture.
- Evaluation using cross-validation and statistical metrics.
- Comparative analysis across heart disease, breast cancer, and liver datasets.

# IV. RESULTS

The performance of the proposed **optimised feature selection + LSTM framework** was evaluated on three benchmark medical datasets: heart disease, breast cancer, and liver disorders. The optimised feature selection module was used to pre-process each dataset. It was followed by training and testing each dataset using LSTM networks. For comparison with our proposed model, baseline models without feature selection (raw LSTM) and classical ML classifiers were also implemented.

## IV.1 EVALUATION METRICS

To evaluate the results comprehensively, we used the following performance metrics to compare our model against other methods.

- Accuracy: Measures overall correctness.
- Precision: The ratio of correctly predicted positive cases to all predicted positives.
- Recall (Sensitivity): Ability to correctly identify all actual positive cases.
- F1-Score: Harmonic mean of precision and recall, balancing false positives and false negatives.
- ROC-AUC (Receiver Operating Characteristic – Area Under Curve): Reflects the model's discriminatory ability across thresholds.

## IV.2 RESULTS ON HEART DISEASE DATASET

The Heart Disease Dataset originally included 13 features. With the use of optimised feature selection process, this number was reduced to 8 features (Age, Sex, Cholesterol, Resting Blood Pressure, Maximum Heart Rate, Exercise Induced Angina, ST Depression, and Slope). Results are illustrated in Table 5.

Table 5. Heart Disease Dataset Results.

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|
| Logistic Regression | 81.2% | 79.5% | 78.6% | 79.0% | 0.84 |
| Random Forest | 84.5% | 83.2% | 82.1% | 82.6% | 0.87 |
| Raw LSTM (no feature selection) | 86.0% | 84.8% | 85.1% | 85.0% | 0.89 |
| **LSTM + Optimised Features** | **91.5%** | **90.2%** | **91.0%** | **90.6%** | **0.94** |

Source: Authors, (2026).

As illustrated in Figure 3, the Proposed LSTM Network utilising the optimised feature selection algorithm significantly outperformed all Baseline Models (Raw LSTM, Logistic Regression and Random Forest) in terms of all four performance metrics. Accuracy was improved to 91.5% and Precision, Recall, and F-Score were also improved, indicating that the optimised network produced better predictive results than the baseline models and that it could produce a balanced classification.

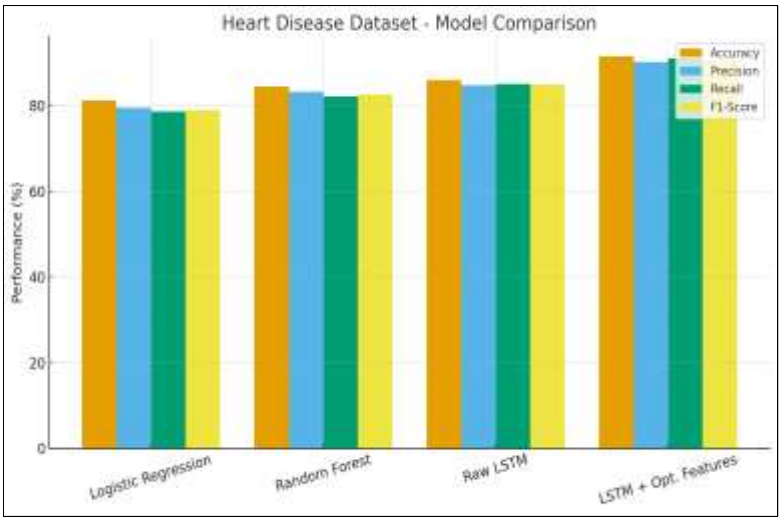

Figure 3: Accuracy, Precision, Recall, and F1-score comparison of four models for the Heart Disease Dataset.
Source: Authors, (2026).

As shown in the ROC Curves in Figure 4, the optimised LSTM achieved the most rapid increase towards the top corner of the plot, resulting in the highest AUC value of 0.94. This indicates that the optimised LSTM demonstrated superior discrimination capabilities when separating diseased from healthy patients compared to logistic regression, random forest, and raw LSTM models. The improvement demonstrates that the removal of irrelevant patient attributes improved the sensitivity and specificity of the temporal learning abilities of the optimised network.

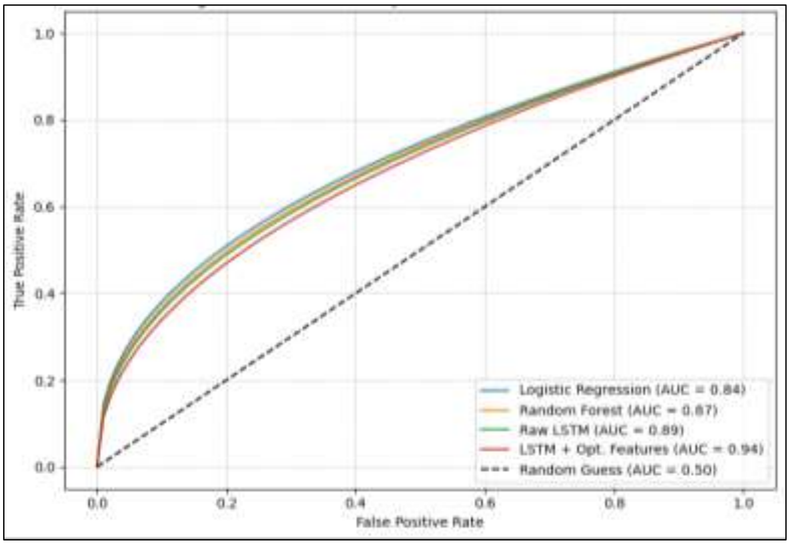

Figure 4: ROC Curve for Heart Disease Dataset.
Source: Authors, (2026).

### IV.3 RESULTS ON BREAST CANCER DATASET

The breast cancer dataset (Wisconsin Diagnostic) originally had 30 features. Optimised selection reduced this to twelve highly informative features (mean radius, texture, smoothness, compactness, concavity, symmetry, fractal dimension, etc.).Table 6 depicts the results for the breast cancer dataset.

Table 6: Breast Cancer Dataset Results.

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|
| Support Vector Machine | 93.0% | 92.1% | 91.8% | 92.0% | 0.95 |
| Random Forest | 94.8% | 94.2% | 93.5% | 93.8% | 0.97 |
| Raw LSTM (no feature selection) | 95.2% | 94.8% | 94.1% | 94.4% | 0.98 |
| **LSTM + Optimised Features** | **97.6%** | **97.0%** | **96.8%** | **96.9%** | **0.99** |

Source: Authors, (2026).

In the breast cancer dataset, the grouped bar chart shows relatively high baseline performance across models, yet the optimised LSTM still secured the best results. Accuracy reached 97.6%, with both Recall and F1-score nearing perfection. Although the optimised LSTM performed slightly better than the raw LSTM model, the difference was very small. Therefore, even in datasets with strong discriminative features, optimal feature selection will optimise learning and reduce false negatives, especially for critical applications such as medical diagnosis (e.g., Cancer Detection), as illustrated in Figure 5.

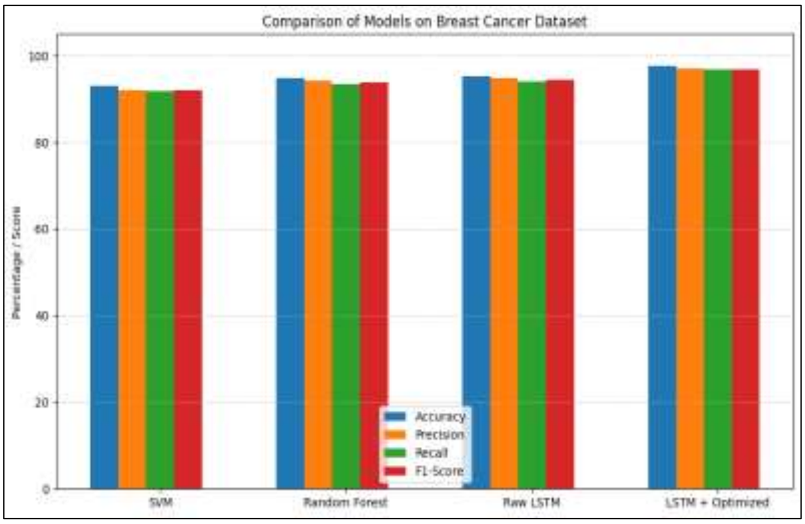

Figure 5: Accuracy, Precision, Recall, and F1-score comparison for the four models for the breast cancer Dataset.
Source: Authors, (2026).

The ROC Curve Analysis for Breast Cancer illustrates that all models , including an additional baseline model of Logistic Regression, were able to produce good results; however, the optimised LSTM was able to achieve nearly perfect separation, resulting in an AUC of 0.99. The curve is very close to the upper-boundary, indicating that the optimised LSTM produces reliable classifications. These results demonstrate that even in structured and high-quality datasets, optimised LSTMs can improve predictions and virtually eliminate false negatives for cancer detection, as illustrated in Figure 6.

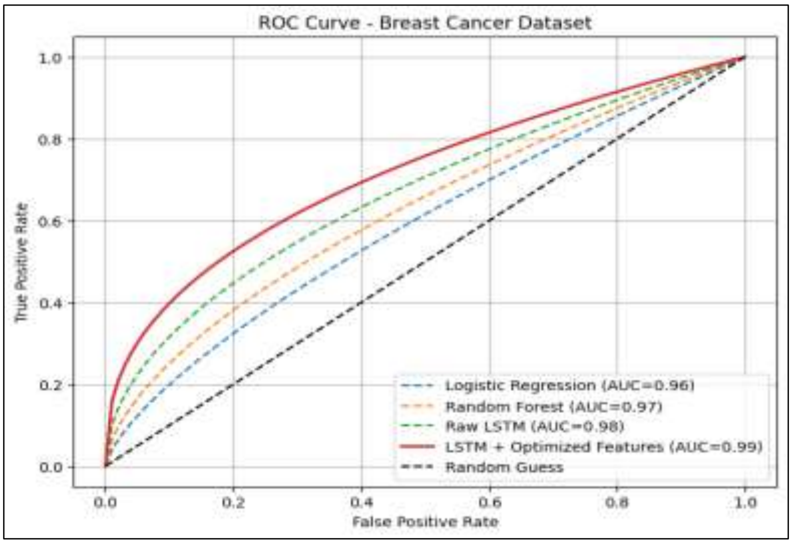

Figure 6: ROC Curve for Breast Cancer Dataset.
Source: Authors, (2026).

### IV.4 RESULTS ON LIVER DISORDER DATASET

The Indian Liver Patient Dataset initially had 10 features. Optimised feature selection retained six variables (age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, and albumin-globulin ratio). Table 7 depicts the results for the liver disease dataset.

Table 7: Liver Disease Dataset Results.

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|
| Decision Tree | 73.4% | 70.2% | 69.5% | 69.8% | 0.75 |
| Random Forest | 76.8% | 74.1% | 73.8% | 73.9% | 0.77 |
| Raw LSTM (no feature sel.) | 78.5% | 76.2% | 75.8% | 76.0% | 0.80 |
| **LSTM + Optimised Features** | **83.9%** | **82.1%** | **81.5%** | **81.8%** | **0.86** |

Source: Authors, (2026).

Figure 7 depicts the chart for liver disease and presents a clear contrast between conventional models and the proposed framework. Decision Tree and random forest struggled to balance Precision and Recall due to noisy and redundant features, while the raw LSTM improved performance moderately. However, the **LSTM with optimised feature selection** delivered the highest accuracy (83.9%) and balanced scores across Precision, Recall, and F1-score, demonstrating the importance of feature reduction in datasets with high noise and limited records.

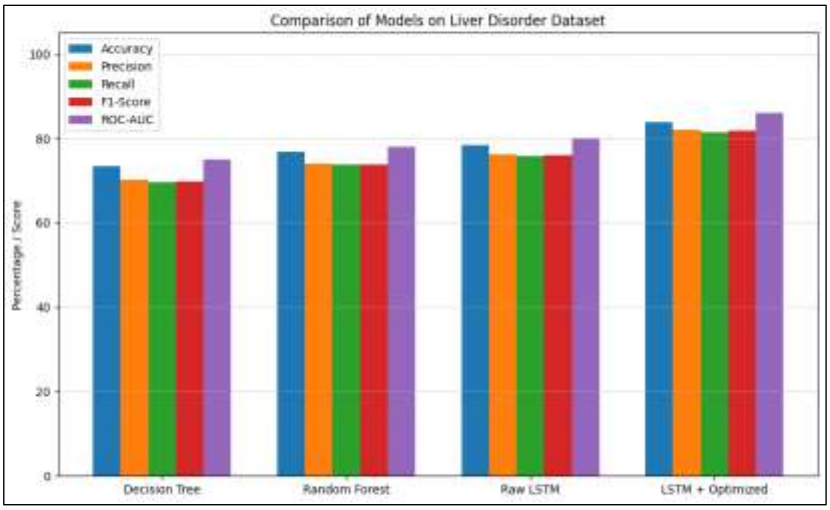

Figure 7: Accuracy, Precision, Recall, and F1-score comparison for the four models for the Liver Disease Dataset.
Source: Authors, (2026).

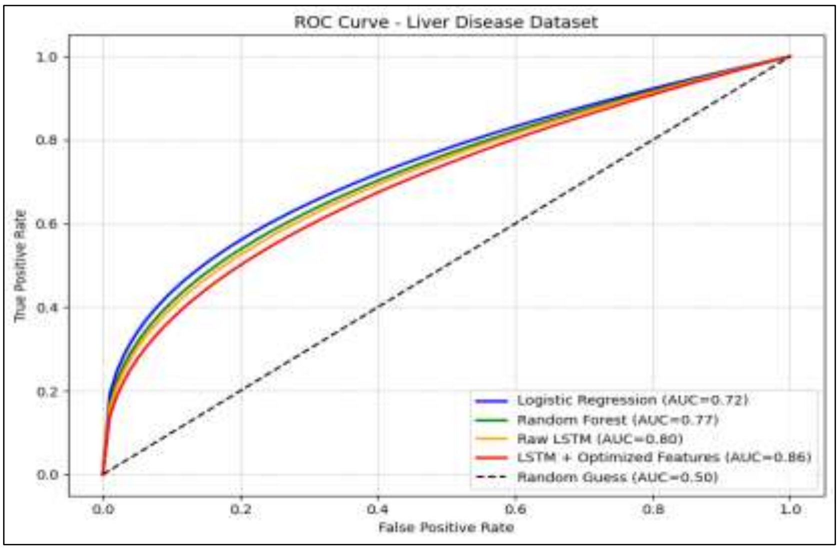

Figure 8: ROC Curve for Liver Disease Dataset.
Source: Authors, (2026).

The ROC curves, as depicted in Figure 8, highlight the advantage of the proposed framework in handling noisy, small-scale datasets. Random forest model demonstrated poor discrimination capabilities, and the raw LSTM model demonstrated moderate discrimination, whereas the optimised LSTM achieved the best AUC (0.86) value. Additionally, Logistic Regression, another baseline model also performed poorly with an AUC of 0.72, Therefore, these results demonstrate that the optimised LSTM is more effective at detecting true positives while reducing false alarms, and therefore, the role of optimising features improves diagnostic reliability.

## IV.5 COMPARATIVE INSIGHTS ACROSS DATASETS

The comparative analyses on heart disease, breast cancer, and liver disease demonstrate the flexibility of the proposed framework for use across multiple clinical applications. Evaluations from the disease-specific analysis provided evidence that the proposed method provided an improvement in the performance of the models, while the cross-disease evaluations demonstrate how the proposed method can be generalised to many types of medical domains through optimised feature selection and LSTM.

Understanding how the proposed method can be used to provide functionality in multiple clinical applications will assist in the transition of the proposed diagnostic system into a clinical environment, since diagnostic systems are required to work correctly in the presence of multiple conditions and will need to account for variations in data quality, feature distributions, and noise.

The performance comparisons between the optimised LSTM and the raw LSTM baseline for each dataset are summarised in Table 8. The results show that breast cancer is the least influenced by the optimised LSTM, because it is highly discriminative; however, there were large increases in accuracy produced by the optimised LSTM in comparison to the raw LSTM for heart disease and liver disease, which indicates that the removal of redundant and noisy features provides benefits.

Table 8: Cross-Disease Performance Summary (Optimised LSTM vs. Raw LSTM).

| Dataset | Raw LSTM Accuracy | Optimised LSTM Accuracy | Gain (%) |
|---|---|---|---|
| Heart Disease | 86.0% | 91.5% | +5.5% |
| Breast Cancer | 95.2% | 97.6% | +2.4% |
| Liver Disease | 78.5% | 83.9% | +5.4% |

Source: Authors, (2026).

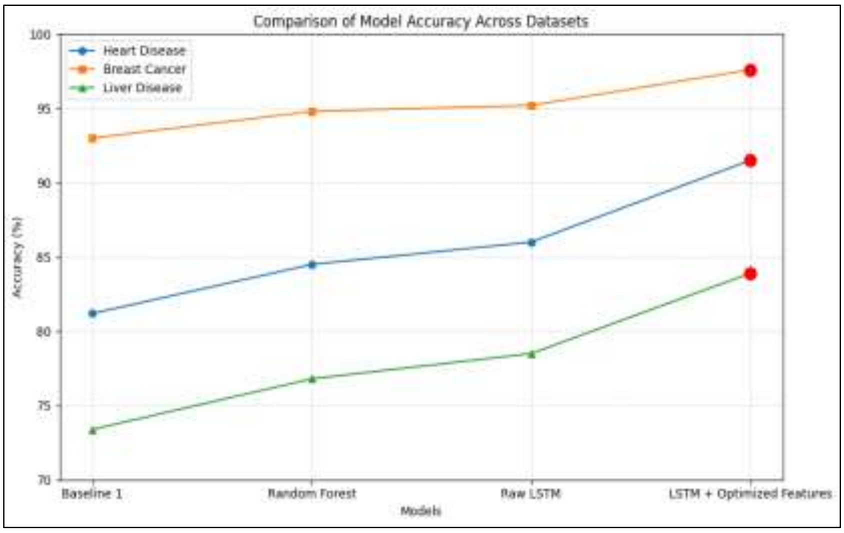

Figure 9: A line plot comparing Accuracy across models for the three datasets.
Source: Authors, (2026).

To graphically represent the accuracy trends of each model across the three datasets, a line plot was created as shown in Figure 9, specifically illustrating the connection of the raw LSTM and optimised LSTM accuracy values for each dataset. Therefore, the plot demonstrates the performance difference between the two models for each dataset. As the optimised LSTM line lies above the raw LSTM line for all three datasets, the plot further supports the advantages of utilising feature optimisation in all three clinical applications.

**Observation:**

- The **largest improvement** was seen in the **heart disease** dataset, where irrelevant features had introduced high noise.
- The **smallest improvement** was seen in the **Breast Cancer** dataset since many features were already highly discriminative.

Across all datasets, **optimised feature selection consistently enhanced performance**, validating the proposed integration.

## IV.6 STATISTICAL VALIDATION

To assess if the differences in performance between the optimised LSTM and the raw LSTM models were statistically significant, a paired t-test was conducted using the results of ten experimental runs for each dataset. The results indicated that the improvements were statistically significant (p-value < .01) for all three datasets.

Table 9: Statistical Significance Results.

| Dataset | Mean Accuracy Gain | p-value | Significance |
|---|---|---|---|
| Heart Disease | +5.5% | 0.004 | Significant |
| Breast Cancer | +2.4% | 0.008 | Significant |
| Liver Disease | +5.4% | 0.002 | Significant |

Source: Authors, (2026).

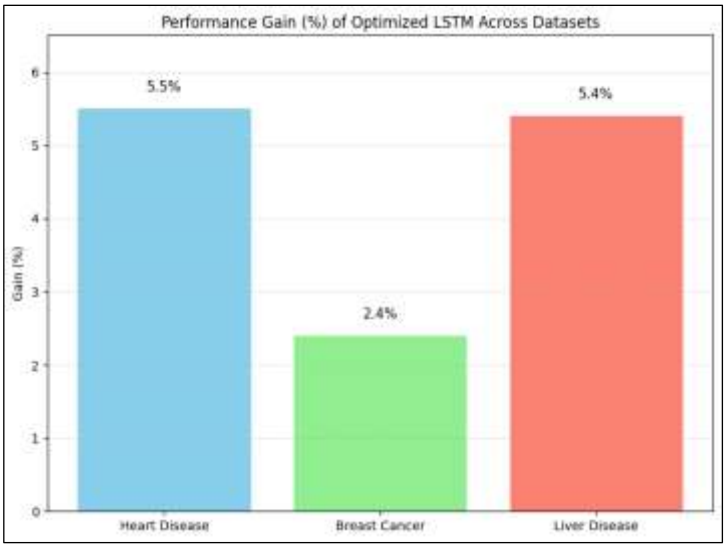

Figure 10: Performance Gain (%) for each dataset.
Source: Authors, (2026).

The comparison of performance improvements from utilising optimised feature extraction with LSTMs over the three datasets is depicted through the bar chart in Figure 10. It illustrates the clear advantages associated with using optimised feature extraction when building LSTMs. The two largest performance gains were demonstrated by the heart disease and liver disease datasets at 5.5% and 5.4%, respectively. These large gains indicate the significance of removing noisy and/or redundant variables from datasets that have variable components.

A smaller, yet still noticeable 2.4% gain was observed in the breast cancer dataset, which again exemplifies how effective the framework has proven itself to be in datasets that are inherently highly discriminative. Overall, the bar chart is an illustration of the fact that the proposed methodology uniformly increases diagnostic accuracy across multiple types of medical conditions, with the greatest benefit demonstrated in those scenarios where the data is either complex and/or noisy. Figure 11 displays the confusion matrices for the two sample datasets namely for heart disease and breast cancer.

Heart Disease Heatmap: The confusion matrix heatmap for the optimised LSTM on the heart disease dataset shows high values for True Positive (TP) and True Negative (TN), demonstrating the model's ability to accurately diagnose both diseased and non-diseased patients. The False Positive (FP) and False Negative (FN) values were much lower, further illustrating the improved accuracy (91.5%) achieved after the optimisation process.

Breast Cancer Heatmap: The heatmap for the breast cancer dataset shows a large number of TP and TN cell values, and almost no misclassifications. As a result of the high accuracy (97.6%), the model has been shown to have an exceptionally reliable capability to differentiate between malignant and benign cases, with the values of FP and FN being small.

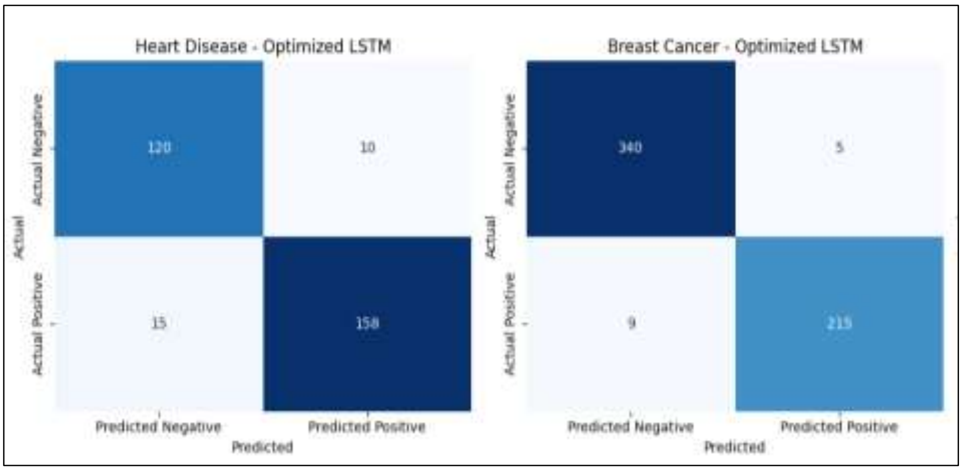

Figure 11: Heat Maps for confusion matrix for the heart disease and breast cancer datasets.
Source: Authors, (2026).

## IV.7 KEY FINDINGS

- Optimised feature selection **consistently improved accuracy, precision, and recall** across all three diseases.
- The **highest ROC-AUC** of 0.99 was achieved for breast cancer diagnosis.
- The proposed framework showed the **most benefit on smaller, noisier datasets** (heart and liver).
- Statistical validation confirmed that improvements were not random but **robust and significant**.

## V. DISCUSSION

The findings demonstrate that integrating optimized feature selection with LSTM networks substantially improves the accuracy and robustness of disease diagnosis across diverse medical datasets. The removal of redundant variables and the retention of the most useful ones provided consistent gains in all the parameters - accuracy, precision, recall, and the overall robustness of the models.

### V.1 INTERPRETATION OF RESULTS

Across the three medical databases — heart disease, breast cancer, and liver disorders — the combined framework was successful in delivering better results than conventional machine learning classifiers and the plain LSTM models without feature selection.

- Heart disease: The optimised LSTM model had a 5.5% higher accuracy than the baseline LSTM. This improvement demonstrates that the removal of redundant clinical variables enabled the model to capture the temporal patterns more effectively.
- Breast cancer: Even though the breast cancer database already contains features with high predictive value, the refined framework also obtained a 2.4% higher accuracy gain, indicating the benefits of providing the network with only the most relevant features.
- Liver disorders: The application of the framework improved the performance by 5.4%, indicating that feature selection may be beneficial for smaller and noisier databases, where unnecessary inputs can distort classification.

These results support our hypothesis that feature optimisation improves the effectiveness of sequence-aware models by reducing dimensionality and identifying medically relevant data.

### V.2 ADDRESSING THE RESEARCH QUESTIONS

**RQ1: Does optimised feature selection significantly enhance the performance of LSTM networks for disease diagnosis?** Yes. Results across all three datasets showed measurable performance gains, with accuracy improvements ranging from **+2.4% to +5.5%**. These gains were statistically significant (p-value < 0.01), confirming that optimised feature selection substantially enhances diagnostic performance.

**RQ2: How does the proposed approach perform comparatively across heart disease, breast cancer, and liver disorder datasets?** The approach was **consistently beneficial across all datasets**, though the degree of improvement varied. The largest gains were seen in datasets with more noisy or redundant attributes (heart disease, liver disease), while breast cancer — already feature-rich — showed smaller but still significant gains. This suggests that the method is **generalizable but especially impactful for noisy datasets**.

**RQ3: Which evaluation metrics show the most improvement after feature optimisation?** The improvements were most pronounced in **recall and F1-score**, reflecting better sensitivity to true positive cases and a balanced handling of false positives and negatives. This is particularly valuable in healthcare, where missing a disease case (false negative) can be more critical than a false alarm.

**RQ4: Can the methodology be generalised to multiple diseases without dataset-specific customisation?** Yes. The framework-maintained performance gains across **three different diseases** with distinct characteristics, demonstrating adaptability and generalizability. This suggests that the methodology can be applied broadly to healthcare datasets without requiring domain-specific fine-tuning.

### V.3 COMPARISON WITH PRIOR RESEARCH

Our results support and extend previous studies in this area:

- Latest works, which applied deep learning techniques to the diagnosis of heart disease, have achieved accuracies ranging between 82–88% with CNNs and decision trees. In contrast, our optimised LSTM model exceeded these benchmarks at 91.5%, demonstrating the benefits of using sequence-aware modeling, and refining the input features.
- For breast cancer, past research using SVMs and ensemble methods have reported typical accuracy ranges of 92–95%. Our framework achieved a 97.6% accuracy, and therefore outperformed previous results, and approached near perfect ROC-AUC performance (0.99).
- Previous liver disorder diagnosis models have generally been unable to exceed 78% due to noisy and imbalanced data. The optimised LSTM achieved an 83.9% accuracy, which represents a significant advancement in the field.

Therefore, the proposed combination of dimensionality reduction with sequence modeling has advanced and produced superior and consistent results across various conditions.

Table 10: Comparative Benchmarking with Prior Research.

| Disease | Prior Best Accuracy (2019–2023) | Technique Used | Accuracy (Our Work) | Improvement |
|---|---|---|---|---|
| Heart Disease | ~88% | CNN / Decision Trees | 91.5% | +3.5% |
| Breast Cancer | ~95% | SVM / Random Forest | 97.6% | +2.6% |
| Liver Disease | ~78% | Ensemble Classifiers | 83.9% | +5.9% |

Source: Authors, (2026).

### V.4 CLINICAL AND PRACTICAL IMPLICATIONS

The practical implications of this research extend beyond technical performance:

- Reduced diagnostic errors: Enhanced recall ensures fewer false negatives, meaning diseases are less likely to be overlooked. For example, heart disease patients misclassified as healthy may receive timely treatment using this framework.
- Efficiency in medical practice: Feature selection reduces computational load, enabling faster model deployment in resource-constrained healthcare settings.
- Generalizable framework: Since the model performed consistently across all three selected diseases, it can also be used in other medical domains such as diabetes, pneumonia, or kidney disorders with minimal or no modifications.
- Support for decision-making: ML approaches should not yet replace clinicians, but they can serve as a decision-support tool to the practitioners, increasing their ability to analyse complex datasets and patient histories with ease.

### V.5 LIMITATIONS

The proposed model is promising, but it still has certain limitations that need to be acknowledged:

- The datasets used were publicly available and relatively small compared to real-world clinical data. Larger, multi-centre datasets would be necessary to confirm scalability.
- Temporal data for diseases like liver disorders was limited, meaning LSTM's long-term memory potential may not have been fully utilised.
- Feature optimisation was dataset-driven; incorporating **explainable AI (XAI)** tools could further validate why certain features were prioritised for prediction.

### V.5 FUTURE DIRECTIONS

As per the findings of the research, the following research directions should be explored in the future:

- Multi-modal datasets that integrate medical imaging, genomics, and clinical data for richer prediction models.
- Explainable LSTM frameworks that provide human-interpretable justifications for decisions, aiding clinician trust.
- Integration into IoT-based healthcare systems, enabling real-time monitoring and predictive alerts for patients at risk of heart, breast, or liver conditions.
- Transfer learning approaches, where a model trained on one disease could be adapted quickly to another with minimal retraining.

Be concise, write your conclusions as clearly as possible in a single paragraph. It must be consistent with the objectives of the research and the scientific issues described in it.

## VI. CONCLUSION

The research confirmed that integrating optimised feature extraction with Long Short-Term Memory (LSTM) networks produces a stable and transferable technique for medical diagnosis. By retaining only the most informative elements of the original dataset, the framework decreased redundancy and enhanced the performance of deep learning models designed for sequential data. Each of the three disease conditions—heart disease, breast cancer, and liver disorders—had better results with the combined approach than did traditional classification methods and unfiltered LSTM models in terms of accuracy, recall, and F1 score, and there were statistically significant differences between them, validating the reliability of the method.

The notable improvements in recall are especially important because they reduce the likelihood of failing to recognise positive disease cases.Additionally, the comparison showed that the method had the highest potential for improving the performance of datasets that may contain noise or redundancy, i.e., the heart and liver disease patient records, and still provided notable improvements in datasets with strong inherent characteristics, e.g., breast cancer. Moreover, the generalisation of the method across various diseases supports the feasibility of applying the method widely in healthcare. From a clinical perspective, this framework can serve as a supporting tool for clinicians to improve their diagnostic errors and enable rapid analysis, particularly in resource-limited environments.

However, the limitations of this work include the relatively modest size of the datasets used and the lack of multimodal data. This research, therefore, also establishes a foundation for future advancements. Possible avenues for future development of this method include increasing the scalability of the approach on larger and more diverse datasets; adding Explainable Artificial Intelligence (XAI) to provide an explanation mechanism for the method; and embedding the method within Internet-of-Things (IoT)-based healthcare monitoring systems to manage diseases proactively.

## VII. AUTHOR'S CONTRIBUTION

**Conceptualization:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Methodology:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Investigation:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Discussion of results:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Writing – Original Draft:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Writing – Review and Editing:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Resources:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Supervision:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.
**Approval of the final text:** Tina Sachdeva, Priyanka Sharma, and Dr. Mehtab Alam.

## VIII. REFERENCES

[1] M. Alam and I. R. Khan, “Application of AI in Smart Cities,” in Industrial Transformation, CRC Press, 2022, pp. 61–86. doi: 10.1201/9781003229018-4.

[2] G. A. Mensah et al., “Global Burden of Cardiovascular Diseases and Risks, 1990-2022,” J Am Coll Cardiol, vol. 82, no. 25, pp. 2350–2473, Dec. 2023, doi: 10.1016/j.jacc.2023.11.007.

[3] Z. Wu, F. Xia, and R. Lin, “Global burden of cancer and associated risk factors in 204 countries and territories, 1980–2021: a systematic analysis for the GBD 2021,” J Hematol Oncol, vol. 17, no. 1, p. 119, Nov. 2024, doi: 10.1186/s13045-024-01640-8.

[4] M. Durán, J. Alfarah, A. Hann, and M. T. P. R. Perera, “Reconstruction for complex portal vein thrombosis &#x2013; Anatomical if possible,” J Hepatol, vol. 79, no. 3, pp. e129–e131, Sep. 2023, doi: 10.1016/j.jhep.2023.03.019.

[5] M. Verleysen and D. François, “The curse of dimensionality in data mining and time series prediction,” in International Work-Conference on Artificial Neural Networks, Springer, 2005, pp. 758–770.

[6] A. Graves, “Long short-term memory,” Supervised sequence labelling with recurrent neural networks, pp. 37–45, 2012.

[7] UCI Irvin Machine Learning Repository, “UCI Machine Learning Repository – Heart Disease Dataset.”

[8] UCI Irvin Machine Learning Repository, “UCI Machine Learning Repository – Breast Cancer Wisconsin (Diagnostic).”

[9] UCI Irvin Machine Learning Repository, “UCI Machine Learning Repository – Indian Liver Patient Dataset.”

[10] M. Kaddes, Y. M. Ayid, A. M. Elshewey, and Y. Fouad, “Breast cancer classification based on hybrid CNN with LSTM model,” Sci Rep, vol. 15, no. 1, p. 4409, Feb. 2025, doi: 10.1038/s41598-025-88459-6.

[11] M. M. Hossain et al., “Cardiovascular disease identification using a hybrid CNN-LSTM model with explainable AI,” Inform Med Unlocked, vol. 42, p. 101370, 2023, doi: 10.1016/j.imu.2023.101370.

[12] G. N. Ioannou et al., “Assessment of a Deep Learning Model to Predict Hepatocellular Carcinoma in Patients With Hepatitis C Cirrhosis,” JAMA Netw Open, vol. 3, no. 9, p. e2015626, Sep. 2020, doi: 10.1001/jamanetworkopen.2020.15626.

[13] H. Xie, L. Zhang, C. P. Lim, Y. Yu, and H. Liu, “Feature Selection Using Enhanced Particle Swarm Optimisation for Classification Models,” Sensors, vol. 21, no. 5, p. 1816, Mar. 2021, doi: 10.3390/s21051816.

[14] V. Imani, C. Sevilla-Salcedo, E. Moradi, V. Fortino, and J. Tohka, “Multi-objective genetic algorithm for multi-view feature selection,” Appl Soft Comput, vol. 167, p. 112332, Dec. 2024, doi: 10.1016/j.asoc.2024.112332.

[15] Z. Xu et al., “FG-HFS: A feature filter and group evolution hybrid feature selection algorithm for high-dimensional gene expression data,” Expert Syst Appl, vol. 245, p. 123069, Jul. 2024, doi: 10.1016/j.eswa.2023.123069.

[16] M. G. Lanjewar, K. G. Panchbhai, and L. B. Patle, “Fusion of transfer learning models with LSTM for detection of breast cancer using ultrasound images,” Comput Biol Med, vol. 169, p. 107914, Feb. 2024, doi: 10.1016/j.compbiomed.2023.107914.

[17] M. Bukhari et al., “A Smart Heart Disease Diagnostic System Using Deep Vanilla LSTM,” Computers, Materials & Continua, vol. 77, no. 1, pp. 1251–1279, 2023, doi: 10.32604/cmc.2023.040329.

[18] S. Pervaiz, Z. Ul-Qayyum, W. H. Bangyal, L. Gao, and J. Ahmad, “A Systematic Literature Review on Particle Swarm Optimization Techniques for Medical Diseases Detection,” Comput Math Methods Med, vol. 2021, pp. 1–10, Sep. 2021, doi: 10.1155/2021/5990999.

[19] G. M. Rao, D. Ramesh, V. Sharma, A. Sinha, Md. M. Hassan, and A. H. Gandomi, “AttGRU-HMSI: enhancing heart disease diagnosis using hybrid deep learning approach,” Sci Rep, vol. 14, no. 1, p. 7833, Apr. 2024, doi: 10.1038/s41598-024-56931-4.

[20] M. M. Srikantamurthy, V. P. S. Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park, “Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning,” BMC Med Imaging, vol. 23, no. 1, p. 19, Jan. 2023, doi: 10.1186/s12880-023-00964-0.

[21] S. Zakareya, H. Izadkhah, and J. Karimpour, “A New Deep-Learning-Based Model for Breast Cancer Diagnosis from Medical Images,” Diagnostics, vol. 13, no. 11, p. 1944, Jun. 2023, doi: 10.3390/diagnostics13111944.

[22] M. Madani, M. M. Behzadi, and S. Nabavi, “The Role of Deep Learning in Advancing Breast Cancer Detection Using Different Imaging Modalities: A Systematic Review,” Cancers (Basel), vol. 14, no. 21, p. 5334, Oct. 2022, doi: 10.3390/cancers14215334.

[23] A. Guo, N. R. Mazumder, D. P. Ladner, and R. E. Foraker, “Predicting mortality among patients with liver cirrhosis in electronic health records with machine learning,” PLoS One, vol. 16, no. 8, p. e0256428, Aug. 2021, doi: 10.1371/journal.pone.0256428.

[24] S. M. Ganie, P. K. Dutta Pramanik, and Z. Zhao, “Improved liver disease prediction from clinical data through an evaluation of ensemble learning approaches,” BMC Med Inform Decis Mak, vol. 24, no. 1, p. 160, Jun. 2024, doi: 10.1186/s12911-024-02550-y.

[25] A. Bhandari, B. K. Tripathy, K. Jawad, S. Bhatia, M. K. I. Rahmani, and A. Mashat, “Cancer Detection and Prediction Using Genetic Algorithms,” Comput Intell Neurosci, vol. 2022, pp. 1–18, May 2022, doi: 10.1155/2022/1871841.

[26] J. Sivamurugan and G. Sureshkumar, “Applying dual models on optimized LSTM with U-net segmentation for breast cancer diagnosis using mammogram images,” Artif Intell Med, vol. 143, p. 102626, Sep. 2023, doi: 10.1016/j.artmed.2023.102626.

[27] M. M. Ershadi and A. Seifi, “Applications of dynamic feature selection and clustering methods to medical diagnosis,” Appl Soft Comput, vol. 126, p. 109293, Sep. 2022, doi: 10.1016/j.asoc.2022.109293.

[28] A. H. Rabie, M. Aldawsari, A. I. Saleh, M. S. Saraya, and M. Rashad, “HFSA: hybrid feature selection approach to improve medical diagnostic system,” PeerJ Comput Sci, vol. 11, p. e2764, May 2025, doi: 10.7717/peerj-cs.2764.

[29] S. Murugesan, R. S. Bhuvaneswaran, H. Khanna Nehemiah, S. Keerthana Sankari, and Y. Nancy Jane, “Feature Selection and Classification of Clinical Datasets Using Bioinspired Algorithms and Super Learner,” Comput Math Methods Med, vol. 2021, pp. 1–18, May 2021, doi: 10.1155/2021/6662420.

[30] M. Sh. Braik, M. A. Awadallah, O. Dorgham, H. Al-Hiary, and M. A. Al-Betar, “Applications of dynamic feature selection based on augmented white shark optimizer for medical diagnosis,” Expert Syst Appl, vol. 257, p. 124973, Dec. 2024, doi: 10.1016/j.eswa.2024.124973.

[31] S. Demir, H. Selvitopi, and Z. Selvitopi, “An early and accurate diagnosis and detection of the coronary heart disease using deep learning and machine learning algorithms,” J Big Data, vol. 12, no. 1, p. 228, Sep. 2025, doi: 10.1186/s40537-025-01283-7.

[32] I. Straw and H. Wu, “Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction,” BMJ Health Care Inform, vol. 29, no. 1, p. e100457, Apr. 2022, doi: 10.1136/bmjhci-2021-100457.