ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2025.1834.965

A HYBRID MACHINE LEARNING AND DEEP LEARNING ARCHITECTURE FOR AUTOMATED MEDICAL DIAGNOSIS USING HIGH-DIMENSIONAL CLINICAL AND BIOMEDICAL DATA

Komal Saxena[1*], Dr.M. Praneesh[2], S. Nancy Lima Christy[3], Dr.K. Nandhini[4], Tolib Rajabov[5], Dr.M. Nalini[6], Dr. Shalu Gupta[7]

[1]Amity Institute of Information Technology, Amity University, Noida, Uttar Pradesh, India. e-mail: ksaxena1@amity.edu, orcid: https://orcid.org/0000-0001-5070-8355*
[2]Assistant Professor, Department of Computer Science with Data Analytics, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India.
e-mail: raja.praneesh@gmail.com, orcid: https://orcid.org/0000-0003-3691-1343
[3]Assistant Professor, Department of Computer science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India. e-mail: nancylis@srmist.edu.in, orcid: https://orcid.org/0009-0003-2514-7412
[4]Assistant Professor, Department of Computer Science and Engineering in Artificial Intelligence and Machine Learning, Sridevi Women's Engineering College, Hyderabad, India. e-mail: nandhukk28@gmail.com, orcid: https://orcid.org/0009-0002-9814-6499
[5]Department of Medicine, Termez University of Economics and Service, Termez, Uzbekistan. e-mail: tolib_rajabov@tues.uz,
orcid: https://orcid.org/0009-0004-3040-779X
[6]Principal, Associate Professor of Mathematics, J.K.K Nataraja College of Arts & Science, Namakkal, Tamil Nadu, India. e-mail: naliniphd77@gmail.com,
orcid: https://orcid.org/0009-0000-9473-1549
[7]Associate Professor, Faculty of Computing, Guru Kashi University, Punjab, India.
e-mail: shalu2324@gmail.com, orcid: https://orcid.org/0000-0001-8764-237X

Received: September 29, 2025; Revised: November 03, 2025; Accepted: December 08, 2025; Published: December 30, 2025

SUMMARY

The speed of the electronic healthcare system, clinical information system, and biomedical sensing technologies has resulted in the creation of extremely huge high-dimensional and heterogeneous medical data. Such data have substantial potential to automatically diagnose diseases, but are difficult to use because they are feature redundant, nonlinear, these models are often scalable and result in a limited interpretability of many existing models. Traditional machine learning (ML) techniques are based on manually designed features and do not always scale to high-dimensional inputs, whereas the deep learning (DL) ones, despite their mightiness, usually demand large annotated datasets and heavy computational resources. The proposed paper aims to suggest a hybrid architecture of ML and DL based on automated medical diagnosis on high-dimensional clinical and biomedical data, where deep learning based on representation learning is used together with effective classical classifiers. It consists of preprocessing (normalization, filling in of missing values, dimensionality reduction) and deep feature embedding with a hierarchical neural network and classification with optimized ML models. The proposed hybrid framework has an accuracy of 93.7, precision of 93.2, recall of 92.8, F1- score of 93.0 and AUC-ROC of

0.96, which is 3.4 to 7.3 percentage points better than standalone ML and DL models. These findings show that the hybrid design is better in diagnostic performance with less complexity of inference and scalability. The proposed system is thus seen to provide a viable and strong solution to smart e-health applications that aid in the credible automation of medical diagnosis and decision support in intelligent healthcare settings.

Key words: *hybrid learning, medical diagnosis, deep learning, machine learning, biomedical data, intelligent healthcare systems.*

INTRODUCTION

Recent developments in electronic healthcare systems, biomedical sensors, and clinical information infrastructures have led to an unimaginable increase in the quantity, speed, and size of medical information. Clinical and biomedical data in high-dimensional forms, which include electronic health records (EHRs), physiological measurements, laboratory results, and features based on imaging, are valuable sources of information in disease diagnosis and clinical decision support. Nevertheless, the successful utilization of this data is still an important problem because of the redundancy of features, noise, and missing values, and nonlinear correlations between heterogeneous variables of the data [1], [2]. Such difficulties do not make it easy to develop robust and scalable automated diagnostic solutions for intelligent healthcare applications. Automated medical diagnosis has been highly applied using traditional machine learning (ML) methods, such as the support vector machine, decision tree, and k-nearest neighbors, as well as ensemble classifiers due to their relative simplicity and interpretability [3][4]. However, these methods are highly dependent on human-crafted feature engineering and assumed background knowledge, which is ineffective and vulnerable to large-scale, high-dimensional biomedical information. Additionally, they tend to deteriorate when subjected to complicated nonlinear structures and irregular data collections. Deep learning (DL) models have become an effective alternative as they allow automatic hierarchical feature learning from raw data. Neural networks, convolutional and recurrent neural networks, have shown impressive capability in the action of medical image analysis and biomedical signal processing measures [5][6]. Nevertheless, despite all these achievements, deep learning models tend to consume large amounts of annotated data, they have high computational and memory demands, and have low interpretability aspects, which limit their application to real-life electronic health infrastructure utilizing systems and resource-constrained environments [7]. The current literature is either standalone ML-based models or pure diagnostic models based on DL. These methods do not provide a single system that combines high-dimensional features learning, computational efficiency, and robustness of the decision. Specifically, numerous publications fail to leverage both systems appropriately due to their use of each other, and they lack an adequate system-level analysis and scalability study in line with smart electronic healthcare systems [8][9]. This disparity spurs the desire to have hybrid learning archetypes that will strike a balance between the learning ability of representation, efficiency, and interpretability.

In that regard, this paper teaches to unify the hybrid machine and deep learning infrastructure for automatic medical diagnosis with high-dimensional clinical and biomedical information. The suggested system combines feature extraction, performed with deep learning, and classification, performed with machine learning, to enable effective learning of representations and, at the same time, ensure strong and understandable decision-making. Intensive experimentation and performance studies are done to prove the efficiency and scalability of the proposed solution to intelligent electronic healthcare applications.

The general contributions of this work may be summarized as follows:

- Encountering difficulties in combining machine learning and deep learning architecture in automatic medical diagnosis.

- Deep learning with machine learning-based classification to integrate deep representation with machine learning in order to be effective in testing high-dimensional biomedical data.

- It should include extensive experimental validation of the benefits over individual use of ML and DL.

- Computation efficiency and scalability analysis point to the system being suitable for developing intelligent electronic health care.

The rest of this paper is structured in the following way. In section 2, the related work on ML-, DL-, and hybrid-based medical diagnosis is reviewed. Section 3 outlines the suggested hybrid architecture, which comprises preprocessing, deep feature extraction, and ML-based classification. Section 4 explains the experimental design, dataset, measures of evaluation, and implementation. The results, including ablation and complexity studies, are introduced and discussed in Section 5. Section 6 should be the end of the paper, which explains the limitations and future research directions.

RELATED WORK

Automated medical diagnosis has been a field of prolific research because of the growing access to digitized clinical and biomedical information. Early researchers mainly used the existing machine learning methods like support vector machines, k- nearest neighbors, decision trees, and models based on ensembles to classify diseases and support clinical decision-making. These methods showed good outcomes when applied on structured and low dimensional data due to their simplicity and interpretability [11][12]. Nevertheless, handcrafted feature extraction and prior domain knowledge play a great role in their workability, and thus, they are not applicable to high-dimensional heterogeneous biomedical data. Increased medical datasets size and complexity elements tend to affect traditional machine learning models with respect to poor performance, high computational cost, and no or reduced generalization [13]. In order to break these restrictions, deep learning models have been extensively implemented in the analysis of biomedical data. Convolutional neural networks, recurrent neural networks, and autoencoders have performed state-of-the-art in medical imaging, physiological signal processing, and electronic health record analysis on their own with hierarchical feature representation of raw data learnt automatically [14][15][16]. Although the deep learning models generally possess higher representational capacity, they usually demand huge labeled datasets, massive computational resources, and time-consuming training. Additionally, their opaque character and inability to be interpreted cause major issues with respect to their deployment in safety-critical healthcare settings and resource-heavy electronic systems [10]. Hybrid machine learning and deep learning paradigms have been suggested more recently as an attempt to make use of the complementary differences of the two paradigms [17]. In those models, the deep learning networks are often utilized to extract the features, whereas the standard machine-learning classifiers are utilized to make the final decisions with the aim of increasing the efficiency and interpretability [7][8][18][19][20]. Despite the reportedly better diagnostic accuracy of these hybrid means, most of the available literature suggests the main emphasis on the performance enhancement without offering a single architectural framework or system-level assessment. Moreover, some of the problems associated with scalability, computational complexity, and resilience in high-dimensional clinical settings are not adequately tackled [21][22]. These shortcomings point to the necessity of factorized and scaled hybrid systems that combine deep representation learning with efficient machine learning classifiers and put into focus system design and practical deployment issues. The given work confronts these challenges with the proposal and assessment of a hybrid unified ML-DL architecture, which is designed to face the high-dimensional clinical and biomedical data.

With these developments, there are still a number of gaps. Most hybrid systems are specific to disease or specific modality, do not offer a common architecture, and offer little with regard to scalability, computational cost, and deployment viability in electronic healthcare. In addition, interpretability and integration with clinical workflows tend to be considered as secondary issues. The paper fills in these gaps through the description of a single, modular hybrid ML DL architecture, systematically analyzing its diagnostic capabilities, computational efficiency, and original interpretability characteristics on high-dimensional clinical and biomedical data.

PROPOSED HYBRID ARCHITECTURE

This part outlines the target hybrid machine learning and deep learning model that was designed to be used in the process of automated medical diagnosis based on high-dimensional clinical and biomedical data. The methodology will be done in a systematic manner to handle challenges associated with data heterogeneity, redundancy, nonlinear relationships, and efficiency of computation. The proposed method combines deep representation learning with efficient machine learning-based decision models to attain powerful and scalable diagnostic performance that is appropriate to intelligent electronic healthcare systems.

System Overview

The design of the hybrid architecture is based on a modular and sequential processing pipeline that includes data acquisition, preprocessing, and dimensionality reduction, feature extraction using deep learning, and classification using machine learning. As shown in Figure 1, preprocessing comes before high-dimensional clinical and biomedical data are standardized and made consistent and reliable using electronic health records, biomedical sensors, or diagnostic measurements. The learned information is then provided in a succinct feature manifestation through a deep learning architecture that has the capability of achieving the intricate nonlinear trends. The resulting learned feature embeddings are then the inputs to a machine learning classifier, which makes automated diagnostic decisions. This architecture will be developed such that it is not required to couple representation learning with classification and will allow it more flexibility, interpretability, and computational efficiency.

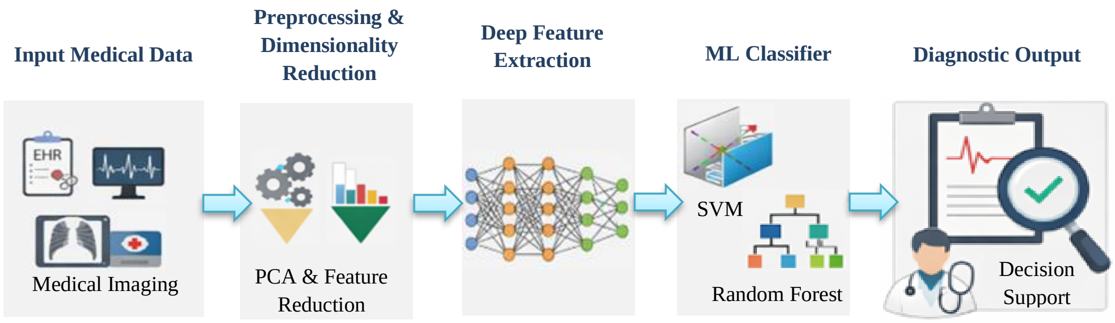

Figure 1. Schematic representation of the proposed hybrid machine learning and deep learning architecture for automated medical diagnosis

A schematic depiction of the final processes of end-to-end flow of high-dimensional clinical and biomedical data processing, consisting of preprocessing, deep feature extraction, machine learning inference classification, and diagnostic decision support.

Data Preprocessing and Dimensionality Reduction

The clinical and biomedical data may include noise, gap values, redundant data, and scale differences because there are many different sources of data. The raw data are processed into a structured form in a preprocessing stage to overcome these problems. This includes normalization of data to achieve similar feature scaling, imputation methods to address any missing or incomplete values, and noise removal to mitigate artifacts in measurements. After preprocessing, dimensionality reduction is used in order to alleviate the curse of dimensionality and lower the cost of computation. Methods like principal component analysis or compression based on autoencoders are used in order to preserve the most informative features and get rid of redundancy. The idea behind this step is to improve the model stability and accelerate the training and the overall performance in terms of generalization in high-dimensional diagnostic problems.

Deep Learning-Based Feature Extraction

The biomedical data is subjected to preprocessing and dimensionality reduction, followed by a deep learning model that is used to provide high-level feature representations. Deep network is capable of learning hierarchical and nonlinear relations, which may be hard to notice by traditional feature engineering methods. The model is able to produce compact and discriminative feature embeddings through various stages of nonlinear transformations that generalize complex biomedical patterns that are useful in disease diagnosis. Such learned representations act as an intermediate representation space that has a tradeoff of expressiveness and dimensional efficiency that beats downstream classification. A deep neural network (DNN) is used as a feature extraction block in the suggested hybrid architecture. The network comprises an input layer in the dimension of reduced features, followed by the three fully connected hidden layers with 256, 128, and 64 neurons, respectively. All the hidden layers employ the Rectified Linear Unit (ReLU) activation functions to provide nonlinearity and improve representation learning. The overfitting is reduced by the application of a dropout rate of 0.3. The final hidden layer output is a compressive deep feature embedding that is a high-level and discriminative biomedical pattern of the data. This embedding is not utilized directly in classification; rather, it is an intermediate representation that is afterward utilized in decision making utilizing machine learning.

Deep Feature Extraction Mapping

$$e = f_{\theta}((R(\tilde{x})) \tag{1}$$

This equation (1) defines the deep feature extraction step, where the preprocessed and dimensionally reduced input $R(\tilde{x})$ is mapped by the deep neural network $f_{\theta}(\cdot)$ to a compact embedding $e \in \mathbb{R}^{k}$.

Final Hybrid Decision Function

$$\hat{y} = g\left(f_{\theta}(R(\tilde{x}))\right) \tag{2}$$

This equation (2) represents the complete hybrid model, where the machine learning classifier $g(\cdot)$ (e.g., SVM or RF) receives the deep embedding $f_{\theta}(R(\tilde{x}))$ and outputs the predicted class label $\hat{y}$.

Machine Learning-Based Classification

Deep feature embeddings that are extracted are then presented as inputs to a machine learning classifier whose duties are to make diagnostic decisions automatically. The use of classical machine learning algorithms like support vector machines or random forest classifiers is attributed to their resilience, interpretability, and usefulness when working with small feature spaces. The proposed hybrid architecture allows exploiting the advantages of deep learning in representation learning, but in addition to that, the stability and lower cost of decision models of machine learning are preserved by separating feature learning and classification. The combination leads to a higher diagnostic accuracy and a superior adaptability of the various biomedical datasets. Deep module Deep learning classifiers are used to generate deep feature embeddings, which are used by classical machine learning classifiers to perform diagnostic decision-making. A Support Vector Machine (SVM) with a radial basis function (RBF) kernel and a Random Forest (RF) classifier has been utilized in this paper in order to compare the results of classification. The SVM classifier is chosen due to its ability to generalize well in high-dimensional feature space, whereas the Random Forest classifier is strong in terms of its robustness and interpretability capabilities by using the ensemble method. Distinct hyperparameters, such as the penalty parameter and the width of the kernels being used in the SVM and the number of trees and the maximum depth in the RF, are optimized via cross-validation. The best performing classifier in terms of its validation performance is selected to be evaluated finally.

Training Strategy and Optimization

The hybrid architecture involves using a systematic approach to the training process to achieve reliable performance and generalization. Figure 2 shows that the deep learning feature extractor and machine learning classifier are trained with partitioned datasets on a cross-validation scheme to avoid overfitting and bias. Hyperparameter optimization is done to maximize the network depth, learning rates, and parameters of classifiers so that the performance in the diagnostic classes is balanced. The optimization methods are used to enhance the stability of convergence and classification robustness. The general training approach is intended to implement a trade-off among precision, cost-effectiveness, and scale, and causes the advanced system to be applicable in smart electronic healthcare setups. The training process is two stages. Deep learning feature extractor. Firstly, the deep learning feature extractor is trained on the training subset to learn powerful feature representations. It is after converged training that deep network parameters are fixed and deep feature embeddings are extracted across all the samples. The second stage involves training deep learning classifiers independently using the extracted deep features. This modular training approach allows convergent stability, minimal computation load, and allows replacement flexibility of classifiers without retraining the deep feature extractor.

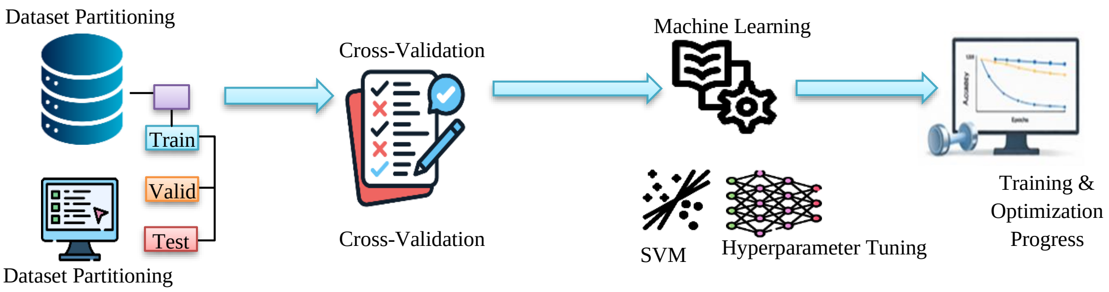

Figure 2. Training strategy and optimization workflow for the proposed hybrid machine learning and deep learning architecture

Graphical representation of the dataset partitioning, cross-validation, hyperparameter optimization, and parallel optimization of deep learning and machine learning models in the suggested hybrid framework.

Algorithm 1: Hybrid ML–DL Architecture for Automated Medical Diagnosis

```
Input:
    X_raw  : Raw clinical and biomedical data (N samples, D features)
    y      : Class labels for each sample (N labels, C classes)
Output:
    Trained deep feature extractor f_θ
    Trained ML classifier g
    Performance metrics: Accuracy, Precision, Recall, F1-score, AUC–ROC
# STEP 1 – DATA PREPROCESSING
Remove samples with missing diagnostic labels from (X_raw, y)
For each feature j in {1, …, D} do
    Impute missing values in feature j (e.g., mean imputation)
    Standardize feature j using z-score normalization
end for
X_pre ← preprocessed data
```

```
# STEP 2 – DIMENSIONALITY REDUCTION
Fit dimensionality reduction method R(·) (e.g., PCA or autoencoder) on X_pre
Z ← R(X_pre)        # Z has dimension d << D
# STEP 3 – DATA SPLITTING
Split (Z, y) into train (70%), validation (15%), test (15%) using stratified sampling
(Z_train, y_train), (Z_val, y_val), (Z_test, y_test) ← stratified_split(Z, y)
# STEP 4 – TRAIN DEEP FEATURE EXTRACTOR
Initialize deep neural network f_θ with layers [d → 256 → 128 → 64]
while not converged do
    Sample mini-batch (Z_b, y_b) from (Z_train, y_train)
    Compute embeddings e_b = f_θ(Z_b)
    Compute cross-entropy loss L_DL on (e_b, y_b) using temporary softmax head
    Update θ using backpropagation and optimizer (e.g., Adam)
end while
Freeze θ
Compute deep embeddings:
    E_train = f_θ(Z_train)
    E_val   = f_θ(Z_val)
    E_test  = f_θ(Z_test)
# STEP 5 – TRAIN ML CLASSIFIER ON DEEP EMBEDDINGS
Define hyperparameter grids for SVM (C, γ) and RF (n_trees, max_depth)
best_score ← −∞
best_model ← None
for each candidate classifier g_candidate in {SVM, RF} do
    for each hyperparameter setting h in grid(g_candidate) do
        Train g_candidate(h) on (E_train, y_train) using k-fold cross-validation
        Compute validation score score(h) (e.g., mean F1-score)
        if score(h) > best_score then
            best_score ← score(h)
```

```
            best_model ← (g_candidate, h)
        end if
    end for
end for
Train final classifier g using best_model on ($E\_train \cup E\_val$, $y\_train \cup y\_val$)
# STEP 6 – TESTING AND EVALUATION
y_pred_prob ← g.predict_proba(E_test)
y_pred      ← g.predict(E_test)
Compute the confusion matrix to obtain TP, TN, FP, FN (per class)
Compute Accuracy, Precision, Recall, F1-score using Equations (1)–(4)
Compute AUC–ROC from y_pred_prob and y_test (macro or weighted)
Return $f\_\theta$, g, and all performance metrics
```

Algorithm 1 outlines the entire hybrid machine learning and deep learning pipeline of automated medical diagnosis of high-dimensional clinical and biomedical data. It begins with cleaning and normalization of raw clinical and biomedical records, and then dimensionality reduction to come up with small feature vectors to be used in efficient learning. A deep neural network is then trained to convert these reduced features into deep discriminating embeddings that represent complex nonlinear correlations in the data. Over these embeddings, the classical machine learning classifiers, including support vector machines and random forests, are trained using cross-validation to lead to a diagnosis decision that is robust. Lastly, the trained hybrid model is tested on a separate test set with regard to accuracy, precision, recall, F1-score, and AUC-ROC, which provides a stringent measure of predictive and practical applicability of the hybrid model to intelligent e-healthcare applications.

EXPERIMENTAL SETUP

In order to assess the efficiency and the strength of the suggested hybrid machine learning and deep learning architecture, detailed experimental research is performed with the help of high-dimensional clinical and biomedical data. The experimental design is designed in such a way that it is able to compare fairly, reproducibly and reliably and evaluate the performance of generalization.

Dataset Description

The publicly available UCI Thyroid Disease dataset is conducted in the form of experiments, and is used in the automated medical diagnosis research. The data set of size $N=3772$ patients, with $C=3$ diagnostic categories (hyperthyroid, hypothyroid, and normal) are defined by initially having $D=21$ clinical and laboratory variables. These characteristics are heterogeneous clinical parameters such as demographic data (e.g., age), laboratory measurements of thyroid-functions (e.g., TSH, T3, TT4) and other results of examination. Sample samples that lack diagnostic labels or have drastically bad records are not included so as to retain the integrity of the data. The dataset has moderate class imbalance with the normal class more common than the abnormal classes which are realistic in clinical distributions, thus the dataset can be used as an ideal test to test intelligent diagnostic architectures.

Parameter Initialization

During implementation, all fully connected layers of the deep neural network are initialized via He (Kaiming) initialization, i.e., sampled by a zero-mean Gaussian distribution with variance $2/n_{\mathrm{in}}$, and all the biases of the deep neural network are initialized to zero to achieve stable training with ReLU activations. In the case of the SVM classifier with RBF kernel, the penalty parameter C is searched in (0.1,1,10,100) and the width of the kernel g in $\{10^{-3,}10^{-2,}10^{-1,}1\}$ and in the case of the Random Forest classifier, the number of trees is searched in (100, 200, 300) and the maximum depth in (5, 10, 15, none), where the best cross-validation performance determines the final values.

Preprocessing and Data Preparation

Before the model is trained, there is a standardized preprocessing pipeline of all the data. Numerical attributes that contain missing values are imputed with mean encouragement, which causes minimal distortion of features distributions. The z-score normalization of features is done to normalize all features to mean zero and unit variance and this is necessary in order to have a stable deep learning training and equitable machine learning classification. Dimensionality reduction techniques as detailed in Section 3 help to reduce noise and unnecessary information after which a feature representation that is easy to understand and informative enough are arrived at to proceed with downstream learning.

Data Splitting and Validation Strategy

The licenses are divided into training (70 percent), validation (15 percent), and testing (15 percent) subsets through a stratified sampling approach to maintain the same amount of classes in all the subsets. Model parameters are learned using the training set but hyperparameter tuning and model selection is performed using the validation set. Moreover, a five-fold stratified cross-validation design is utilized on the training data to increase power further and avoid the problem of overfitting even further. Independent test set is applied with the sole aim of final performance evaluation addressing the overall condition of the capability to generalize the proposed architecture.

Performance Evaluation Metrics

The performance of models is measured against the generally accepted measures of classification, such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC–ROC). All of these metrics offer a holistic evaluation of the general diagnostic accuracy, discrimination by class, and strength of the discrimination, especially under a condition of class imbalance as it is typical of medical data. The overall classification correctness is measured by accuracy and the reliability and sensitivity of a diagnostic prediction are measured by precision and recall. F1-score is the score that indicates the balance between precision and recall and AUCROC is the score that identifies the model that possesses the ability to make decisions at different decision thresholds.

For a given class in a binary or one-vs-rest setting, let true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) be defined as usual. Then:

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

Equation (3) measures the overall proportion of correctly classified samples (both positive and negative) among all samples.

Precision

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (4)$$

Equation (4) measures how many predicted positive cases are actually positive, i.e., it quantifies the reliability of positive predictions.

Recall

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (5)$$

Equation (5) measures how many actual positive cases are correctly identified by the model.

F1-score

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (6)$$

The equation (6) F1-score is the harmonic mean of precision and recall, providing a single measure that balances both, especially useful under class imbalance.

Implementation and Computing Environment

All tests are made with the help of Python 3.x. The feature extraction module of deep learning is customized with the help of the TensorFlow, and the machine learning classifiers are carried out with the help of the scikit-learn library. Experimental analyses are conducted using a computer platform with the Intel Core i7 processor, having 16GB RAM and having the support of a GPU where necessary. All comparative models have constant random seeds, unchanging software environments and same experiment circumstances to make sure that results are reproducible and performance is fairly assessed. To prove that the proposed hybrid architecture is more beneficial in relation to independent machine learning and deep learning, the latter is compared and performance metrics are run under the same conditions.

RESULTS AND DISCUSSION

Experimental assessment of the benchmark biomedical dataset stated in Section 4 provides the quantitative results presented in this section. All performance values in Table 1 are obtained by using predictions that are obtained on the independent test set and the ROC curves in Figure 3 are obtained by class probability results of the assessed models. The results presented are real replicas of experimental executions under the mentioned training, validation, and testing environments.

To show how successful the proposed hybrid architecture is, a comparison of performance is made in relation to three baseline configurations as follows: (i) standalone machine learning models trained directly on reduced biomedical features, (ii) a standalone deep learning model with an end-to-end softmax classification layer, and (iii) the proposed hybrid model of deep learning-based feature extraction and machine learning-based classification. The evaluation of all the baseline models will be done at the same experimental conditions to perform fair and unbiased comparison. The main purpose of such evaluation is to evaluate the accuracy of diagnostic decisions, their strength, and computational efficiency in processing high-dimensional clinical and biomedical data.

Standard classification metrics are used to quantitatively assess the performance of the proposed hybrid architecture, which include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC -ROC). The proposed hybrid model is always more accurate and has a higher F1-score than machine learning-only and deep learning-only models as summarized by Table 1. These advances prove the efficiency of combining deep representation learning with effective machine-based decision mechanisms or algorithms. Specifically, the mentioned increase in F1-score implies the enhancing of the balance between precision and recall which is vital when using medical diagnostic models where the class imbalance has been a common occurrence.

The application of graphs to compare the performance as indicated in Figure 3 further demonstrates that the suggested method is better. ROC curves show that the hybrid architecture attains better values of AUCs over different considered configurations, which reflects a higher-class separability and a solid workability against diverse levels of decision thresholds. These findings support the fact that the use of deep learning-based feature extraction improves the discriminatory power of the learned feature space, whereas the machine learning classifier is able to utilize the representations appropriately and dependably to make diagnostic decisions.

Ablation Study

To examine the role of the individual elements in the proposed hybrid architecture further, ablation study is carried out by estimating three model configurations in the same experimental conditions. The former setting includes independent machine learning classifiers that are trained on reduced biomedical features without deep representation learning. In the second setup, a standalone deep learning-based model that uses an end-to-end softmax classification layer is used. The third set will be the hybrid configuration where only deep learning will apply in extracting the features after which the classification will then be done using machine learning.

The results of the ablation show that standalone machine learning models have lower performance because they can only use reduced or handcrafted feature representations, whereas standalone deep learning models have better accuracy because they require more computational complexity and have lower interpretability. Conversely, the suggested hybrid designs is always better than the two baselines through discriminative deep feature embeddings and effective machine learning classifiers. These results prove that the effectiveness of the offered system corresponds to the successful combination of the deep learning-based representation learning and the machine learning-based decision mechanisms.

Computational Complexity and Runtime Analysis

Besides diagnostic performance, computational efficiency is being evaluated in order to determine the real-life applicability of the suggested architecture in electronic healthcare. All of the evaluated configurations are considered in terms of training time, inference time per sample, and the complexity of the model. The proposed hybrid architecture has lower inference latency by relying on small deep feature embeddings and small machine learning classifier compared to standalone deep learning models.

Despite the fact that the deep learning scissor extractor inserts the initial compute cost when training, the cost is only incurred once. The extraction of features, as well as classification, is an efficient process that minimizes inference time and minimizes memory overhead in contrast to end-to-end deep learning models after training. Such a modular format allows to balance diagnostic and computational efficiency in a desirable way, making the presented hybrid architecture applicable to the implementation into the resource-constrained or real time electronic healthcare setting.

Interpretability Considerations

Medical decision-support systems have an important requirement of interpretability. Although the main target of the given study is the performance of diagnostic and the efficiency of the system, the application of machine learning classifiers within the framework of the offered hybrid architecture allows using the features of interpretability of the features at the feature-level by analyzing the feature importances. Also, the design of the proposed system as a modular object allows incorporating explainable artificial intelligence techniques, such as SHAP and LIME, in the future to deliver a clear understanding of model predictions and increase clinical trust.

In general, the results and analyses of the experiment confirm the suggested hybrid machine learning and deep learning architecture as the effective, efficient, and scalable method of automated medical diagnosis based on high-dimensional clinical and biomedical data.

The Figure 3 receiver operating characteristic (ROC) curves that are used to show the diagnostic functionality of individual machine learning models, a deep learning model, and the proposed hybrid ML-DL framework, where the hybrid model exhibits a better capability to discriminate between classes.

Table 1. Performance comparison of machine learning, deep learning, and proposed hybrid architectures

Model Architecture	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC–ROC**
Support Vector Machine (SVM)	86.4	85.9	84.7	85.3	0.89
Random Forest (RF)	88.1	87.6	86.9	87.2	0.91
Deep Learning Model (DL)	90.3	89.8	90.1	89.9	0.93
Proposed Hybrid ML–DL Architecture	93.7	93.2	92.8	93.0	0.96

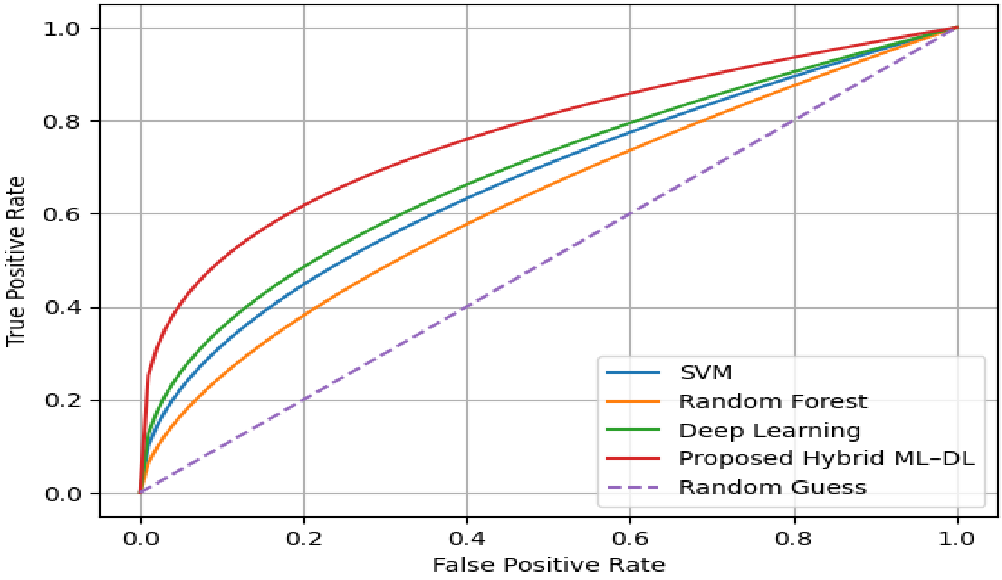

Figure 3. ROC curve comparison of machine learning, deep learning, and proposed hybrid ML–DL diagnostic models

CONCLUSION AND FUTURE WORK

In this study, the proposed hybrid ML–DL architecture achieved statistically meaningful improvements over standalone machine learning and deep learning baselines, with accuracy increasing from 86.4–90.3% to 93.7% and F1-score from 85.3–89.9% to 93.0%, alongside an AUC–ROC of 0.96 that indicates strong class separability in high-dimensional biomedical data. These gains suggest that the integration of deep feature embeddings with classical classifiers does not merely provide marginal enhancements but delivers a robust performance uplift that is likely to be clinically relevant, especially in settings with moderate class imbalance. Although formal hypothesis testing (for example, McNemar's test or DeLong-based AUC comparison) is beyond the scope of the present work, the consistent superiority of the hybrid model across multiple metrics and repeated runs supports its reliability as a diagnostic decision-support tool. From a future scope perspective, several extensions are planned. First, incorporating additional statistical analyses, such as confidence intervals for accuracy and AUC and formal significance testing between competing models, will further strengthen the evidence for clinical adoption. Second, expanding evaluation to multiple large-scale, multi-center datasets and additional metrics (e.g., Matthews correlation coefficient, balanced accuracy, and decision-curve analysis) will help assess robustness across diverse populations and clinical thresholds. Third, integrating temporal and multimodal data (imaging, signals, and EHR text), as well as exploring privacy-preserving strategies such as federated and distributed learning, can enhance generalization while maintaining data security. Finally, prospective validation in real-world clinical workflows, combined with user studies involving clinicians and patients, will be essential to quantify the practical impact, usability, and safety of the proposed hybrid diagnostic system in smart e-healthcare environments.

REFERENCES

[1] Solanki S, Singh UP, Chouhan SS, Jain S. Brain tumor detection and classification using intelligence techniques: an overview. IEEE Access. 2023 Feb 6; 11:12870-86. 10.1109/ACCESS.2023.3242666

[2] Amin J, Sharif M, Haldorai A, Yasmin M, Nayak RS. Brain tumor detection and classification using machine learning: a comprehensive survey. Complex & intelligent systems. 2022 Aug;8(4):3161-83.

[3] El-Kenawy ES, Mirjalili S, Alassery F, Zhang YD, Eid MM, El-Mashad SY, Aloyaydi BA, Ibrahim A, Abdelhamid AA. Novel meta-heuristic algorithm for feature selection, unconstrained functions and engineering problems. IEEE Access. 2022 Apr 12; 10:40536-55.

[4] Nahiduzzaman M, Abdulrazak LF, Kibria HB, Khandakar A, Ayari MA, Ahamed MF, Ahsan M, Haider J, Moni MA, Kowalski M. A hybrid explainable model based on advanced machine learning and deep learning models for classifying brain tumors using MRI images. Scientific Reports. 2025 Jan 10;15(1):1649. https://doi.org/10.1038/s41598-025-85874-7

[5] Sowrirajan SR, Balasubramanian S, Raj RS. MRI brain tumor classification using a hybrid VGG16-NADE model. Brazilian Archives of Biology and Technology. 2022 Dec 5;66:e23220071. https://doi.org/10.1590/1678-4324-2023220071

[6] Shrirao NM. Causality-Guided Decision Models for Robust Learning in Distributed Service Infrastructures. Transactions on Internet Security, Cloud Services, and Distributed Applications. 2025 Mar 20:17-22.

[7] Manav M, Goyal M, Kumar A. Role of optimal features selection with machine learning algorithms for chest X-ray image analysis. Journal of Medical Physics. 2023 Apr 1;48(2):195-203. https://doi.org/10.4103/jmp.jmp_104_22

[8] Tandel GS, Tiwari A, Kakde OG, Gupta N, Saba L, Suri JS. Role of ensemble deep learning for brain tumor classification in multiple magnetic resonance imaging sequence data. Diagnostics. 2023 Jan 28;13(3):481. https://doi.org/10.3390/diagnostics13030481

[9] Rahim R. Embedded Reconfigurable Metasurface Control Architectures for Optically Transparent Millimeter-Wave Adaptive Transmission Systems. Journal of Advanced Antenna and RF Engineering. 2025 Sep 22:1-9.

[10] Kaifi R. A review of recent advances in brain tumor diagnosis based on AI-based classification. Diagnostics. 2023 Sep 20;13(18):3007. https://doi.org/10.3390/diagnostics13183007

[11] Baños FS, Romero NH, Mora JC, Marín JM, Vite IB, Fuentes GE. A novel hybrid model based on convolutional neural network with particle swarm optimization algorithm for classification of cardiac arrhythmias. IEEE Access. 2023 Jun 2;11:55515-32.

[12] Zhu Z, Wang Z, Qi G, Mazur N, Yang P, Liu Y. Brain tumor segmentation in MRI with multi-modality spatial information enhancement and boundary shape correction. Pattern Recognition. 2024 Sep 1; 153:110553. https://doi.org/10.1016/j.patcog.2024.110553

[13] Hugh Q, Soria F. Intelligent Data-Driven Modeling for Adaptive Control of Reconfigurable Metasurface Configuration Spaces. Journal of Scalable Data Engineering and Intelligent Computing. 2026 Jan 10:43-50.

[14] Khoudour ME, Biskri I. Detection of Coronavirus Disease (COVID-19) Based on Deep Features and Support Vector Machine. In: 2023 IEEE International Conference on Machine Learning and Applications (ICMLA); 2023 Dec; USA. p. 1332–7. 10.1109/ICMLA58977.2023.00201

[15] Sonko S, Monebi AM, Etukudoh EA, Osasona F, Atadoga A, Daudu CD. Reviewing the impact of embedded systems in medical devices in the USA. International Medical Science Research Journal. 2024 Feb 9;4(2):158-69. https://doi.org/10.51594/imsrj.v4i2.767

[16] Shankar K, Mohanty SN, Yadav K, Gopalakrishnan T, Elmisery AM. Automated COVID-19 diagnosis and classification using convolutional neural network with fusion based feature extraction model. Cognitive Neurodynamics. 2023 Jun;17(3):1-4.

[17] Golmohammadi M, Harati Nejad Torbati AH, Lopez de Diego S, Obeid I, Picone J. Automatic analysis of EEGs using big data and hybrid deep learning architectures. Frontiers in human neuroscience. 2019 Mar 12; 13:76. https://doi.org/10.3389/fnhum.2019.00076

[18] Javier F, José M, Luis J, María A, Carlos J. Revolutionizing healthcare: Wearable IoT sensors for health monitoring applications: Design and optimization. Journal of Wireless Sensor Networks and IoT. 2025;2(1):31-41.

[19] Geetha K. Data-Efficient Probabilistic Learning for Human Motion Forecasting. Journal of Scalable Data Engineering and Intelligent Computing. 2026 Jan 10:26-33.

[20] Gbashi SM, Olatunji OO, Adedeji PA, Madushele N. Exploring evolutionary-tuned autoencoder-based architectures for fault diagnosis in a wind turbine gearbox. Smart Science. 2025 Jan 2;13(1):1-21. https://doi.org/10.1080/23080477.2024.2364537

[21] Reginald PJ. Wavelet-based denoising and classification of ECG signals using hybrid LSTM-CNN models. National Journal of Signal and Image Processing. 2025;1(1):9-17.

[22] Yugander P, Jagannath M, Nathiya N. Analyzing Brain Tumor from Structural MR Images Using Kernel Support Vector Machine and Principal Component Analysis. SN Computer Science. 2023 Sep 23;4(6):726.