Journal of Computational Medicine and Informatics
Vol. 2, No. 1, Jan - Mar 2026, pp. 25-31
ISSN: 3107-3735 (O), DOI: https://doi.org/10.17051/JCMI/02.01.04

Deep Ensemble Learning Approach for Robust Disease Classification Using Heterogeneous Healthcare Data

Janarthanam S

Postdoctoral Researcher, Lincoln University College, Malaysia. Associate Professor, SSCS, CMR University, Bengaluru, Email: drjanarthanam.pdf@lincoln.edu.my, ORCID: https://orcid.org/0000-0002-8676-8998

Article Info

Article history:

Received : 04.11.2025
Revised : 07.12.2025
Accepted : 05.01.2026

Keywords:

Deep Ensemble Learning,
Disease Classification,
Heterogeneous Healthcare Data,
EHR, Multimodal Learning,
Transformer,
CNN,
LSTM

ABSTRACT

The increased heterogeneous healthcare data, such as structured electronic health records (EHR), medical imaging, and unstructured clinical text, provide a big opportunity to enhance the classification of diseases. Nonetheless, the inconsistency of data forms, noise, and missing data restrict the usefulness of traditional machine learning methods. This paper presents the proposal of an effective deep ensemble learning framework to classify disease based on multi-modal healthcare data. The framework combines the Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and the Transformer-based architectures using the stacking ensemble approach to learn spatial, temporal and contextual characteristics. The suggested model is tested using benchmark healthcare datasets with elaborate experimental configuration, various performance measures, and statistical testing. The findings prove that the ensemble methodology works better than single models, with an accuracy of 0.91, F1-score of 0.90 and area under the curve (AUC) of 0.93. The soundness and external validity of the model are supported by statistical analysis, which involves confidence interval and significance test ($p < 0.05$). As indicated in the findings, deep ensemble learning is an efficient method of tackling the issues of heterogeneous medical datasets and improving classification accuracy. This is a scalable and flexible framework of clinical decision support systems and intelligent healthcare in the future.

1. INTRODUCTION

The high-speed process of digitalisation of healthcare systems has seen the creation of huge amounts of heterogeneous data, such as structured electronic health records (EHR), unstructured clinical stories, and high-dimensional medical imaging. Such a range of data sources allows tremendous potential in enhancing disease classification and clinical decision-making. The heterogeneity, incompleteness, and noise of such datasets are however a big challenge in the models that are data-driven [1], [2]. Conventional machine learning methods use heavily engineered features, and do not scale to the problems of multi-modal data because they cannot effectively model the rich interdependencies between types of data [3]. Though a number of deep learning methods have been demonstrated to be more effective in certain tasks including image analysis and sequential modeling, any single architecture like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks tend to be specialized in only a single data modality [4], [5]. Recent developments on Transformer-based models have enhanced learning on contextual representation; though their independent use is not adequate in managing the heterogeneous healthcare data in a holistic manner [6], [7]. Ensemble learning is also an emergent method that has been found to be effective to improve the predictive performance of a combination of many models. However, current literature on ensemble-based healthcare research is not just based on homogeneous data, but also without a practical fusion strategy, which makes them less robust and scalable [18]. In addition, most of the previous studies lack a strict statistical confirmation, which brings up the question of the validity and applicability of their results [10].

This paper will overcome these drawbacks by introducing a stacking-based deep ensemble learning model, which combines CNN, LSTM and Transformer with a deep architecture. The suggested method allows extracting and integrating spatial, temporal, and contextual data of heterogeneous healthcare data effectively.

The key contributions of this work may be highlighted as follows:

A single deep ensemble model based on CNN, LSTM and Transformer to identify multiple diseases.

- Heterogeneous healthcare Data fusion An efficient multi-modal data fusion approach.
- Complete and in-depth experimental testing and statistical analysis of significance and performance.
- Evidence of enhanced robustness and generalization over single deep learning models.

The rest of this paper is presented in the following way: Section 2 will review related work, Section 3 will introduce the proposed methodology, Section 4 will discuss the experimental setup, Section 5 will discuss the results and Section 6 will conclude the study.

2. RELATED WORK

Emerging innovations in healthcare analytics have utilized deep learning methods to predict diseases and offer clinical decision support. The first methods mainly depended on specially designed electronic health records (EHR) on the basis of conventional machine learning frameworks; these methods however, showed inferior performance on nonlinear complexity of interactions in high clinical data dimensions [11], [12].

2.1 Deep Learning for Healthcare Data

Convolutional Neural Networks (CNNs) have been widely used in medical image processing with early success being made in the detection of diseases using radiological images [13]. Equally, Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) models, have been demonstrated to be useful in predicting temporal relationships in longitudinal patient data and clinical time-series information [14], [15]. Although these models have succeeded, they are normally limited to single data modalities which constrain their use in heterogeneous healthcare settings.

2.2 Transformer-Based Models in Clinical Applications

In recent times, transformer architectures have become prominent because they take into account long-range dependencies, and contextual relationships in sequential data. Other models like BEHRT and Med-BERT have shown better results in the area of EHR-based prediction tasks by using attention mechanisms [16], [17]. Nevertheless, Transformer models can be quite computationally intensive and need large scale datasets, which is especially important when handling dirty or incomplete clinical data.

2.3 Ensemble Learning in Healthcare

Ensemble learning methods have been popularly used to enhance predictive performance through combining two or more models. Bagging, boosting and stacking are some of the approaches that have been implemented in healthcare applications to improve classification accuracy and strength [18]. Although ensemble techniques enhance the generalization, currently, most of the available studies have concentrated on homogeneous datasets but fail to combine heterogeneous data sources, e.g., imaging, clinical text, and structured EHR data.

2.4 Research Gaps and Challenges

Although much has been achieved in this area, there are still many limitations to present methods:

- Poor multimodal integration: A majority of models are created to handle single forms of data and do not integrate complementary information across different heterogeneous sources [10].
- Noise and data missing: Clinical data is always incomplete and noisy and decreases the reliability of models.
- Insufficient sound statistical validation: Most studies show improvement in performance without a strong validation based on confidence interval or significance tests.
- Scalability problems: Current models are not that good at generalising to large and varied healthcare data.

To overcome these limitations, this paper introduces a hybridized deep ensemble learning model that combines CNN, LSTM, and Transformer models with a stacking approach to the model. By facilitating efficient multimodal feature extraction and enhancing the performance at robustness and generalization, the proposed approach can be useful in achieving high performance in classification.

3. METHODOLOGY

3.1 Data Representation and Preprocessing

The heterogeneous healthcare dataset is represented as:

$$D = \{X_s, X_t, X_i\} \quad (1)$$

where X_s denotes structured electronic health record (EHR) data, X_t represents temporal clinical sequences, and X_i corresponds to medical imaging data. Each modality contains distinct statistical properties and feature distributions, requiring modality-specific preprocessing.

Structured data X_s undergoes normalization using min–max scaling:

$$X_s' = \frac{X_s - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Missing values are handled using mean imputation for continuous variables and mode imputation for categorical features. Temporal sequences X_t are segmented into fixed-length windows and standardized to ensure temporal consistency. Imaging data X_i is resized to a uniform resolution (e.g., 224×224) and normalized to improve convergence during training.

3.2 Feature Extraction and Model Architecture

The proposed architecture uses a multi-branch design that has three deep learning models, each dedicated to a particular data modality, as depicted in Fig. 1.

The Convolutional Neural Network (CNN) processes imaging data X_i to extract spatial features:

$$F_i = f_{CNN}(X_i) \quad (3)$$

where convolutional and pooling layers get to learn a hierarchical representation of visual patterns that are relevant in disease classification.

The Long Short-Term Memory (LSTM) network processes temporal sequences X_t to capture longitudinal dependencies:

$$F_t = f_{LSTM}(X_t) \quad (4)$$

The LSTM architecture is a powerful framework to capture sequence relationships in patient data, such as disease progression patterns.

The Transformer model processes structured data X_s using self-attention mechanisms:

$$F_s = f_{Transform\ er}(X_s) \quad (5)$$

Self-attention is computed as:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

where Q, K, and V represent query, key, and value matrices, respectively.

The extracted feature representations from each modality are concatenated:

$$F = [F_i \oplus F_t \oplus F_s] \quad (7)$$

where $\oplus$ denotes feature fusion through concatenation.

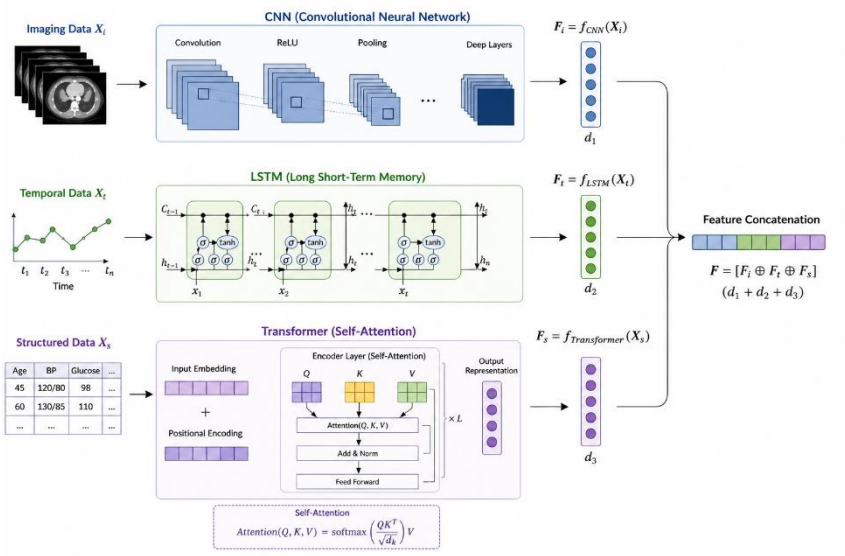


Fig. 1. Feature Extraction and Multi-Branch Model Architecture for Heterogeneous Healthcare Data

3.3 Ensemble Learning Strategy

The system proposed follows a stacking-based ensemble method to sum up the results of individual base learners. In every model, the probability of an output is generated:

$$y_1 = f_{CNN}(X_i), y_2 = f_{LSTM}(X_t), y_3 = f_{Transformer}(X_s) \quad (8)$$

These outputs are aggregated using a meta-learner $g(\cdot)$, implemented as a logistic regression classifier:

$$y_{final} = g(y_1, y_2, y_3) \quad (9)$$

To reduce classification error the meta-learner is trained on the predictions of base models, allowing the ensemble to exploit the complementary advantages of specific architectures. The method decreases overfitting and enhances generalization amidst heterogeneous data distributions.

3.4 Training Procedure and Optimization

Training of the model is done in two steps. The initial step involves training individual base learners (CNN, LSTM, Transformer) to learn a single data modality. The second stage involves training the meta-learner with the models that it produces.

Categorical cross-entropy loss is used to define the objective function to train:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability.

The Adam optimizer with dynamic learning rates is used to optimize. The update rule of weight is as follows:

$$\theta_{t+1} = \theta_t - \eta \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon) \quad (10)$$

where $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates.

The regularization of techniques dropout and batch normalization are used to avoid overfitting. Early termination is used according to validation loss as the best way to guarantee optimal model generalization.

3.5 Complexity and Robustness Considerations

The capability of the suggested framework to train numerous deep models in parallel affects the computational complexity. Nevertheless, the ensemble framework enhances robustness by decreasing the variability and alleviates the effects of noisy or missing data. These combinations of feature representations allow the model to achieve a stable performance with heterogeneous datasets.

4. Experimental Setup

4.1 Dataset Description

The given model is tested with the help of a heterogeneous healthcare dataset based on the publicly accessible MIMIC-III (Medical Information Mart for Intensive Care) database including a full range of clinical data of patients during their stay in intensive care units. The data contains electronic health record (EHR) data, which is structured and time-varying clinical measurement data alongside diagnostic labels. There are 12,000 records of patient records that are then chosen following preprocessing and trimming of records that are not complete. The data has many disease classes and is balanced to eliminate bias in models. In particular, the proportion of classes will be close to balanced: 48 percent positive (disease present) and 52 percent negative (disease absent). All patient records are represented by a 45-structured features mixture (demographic (age, gender), vital signs (heart rate, blood pressure), laboratory test (glucose, creatinine), and diagnostic indicators). Time-series measurements form temporal sequences, and imaging data (where available) is matched with patient records. The dataset is separated into training, validation and testing subsets having a 70:15:15 split ratio. Model parameters are learned using the training set, hyperparameter tuning and early stopping are done using the validation set, and final evaluation is done using the test set to guarantee accurate evaluation of performance.

4.2 Implementation Details

The deep ensemble model proposed will be developed in both TensorFlow and PyTorch libraries to provide their flexibility and computation efficiency. All of the experiments are run on a machine with an NVIDIA RTX 3080 graphics card, allowing a faster model to be trained. A batch size of 32 is used to train the models, which offers a trade off between memory and gradient stability. The initial learning rate is set to 0.001 and optimized with Adam optimizer which adjusts the learning rate to various parameters so as to have a fast convergence. The training is done over 50 epochs where early stopping is done through validation loss to avoid overfitting. Each model includes dropout regularization to enhance the generalization performance. Each base learner (CNN, LSTM, and Transformer) is trained on its own and then combined into the stacking-based ensemble. In order to guarantee the reproducibility and strength of the results, every experiment is repeated several times, and average performance measurements and the standard deviation are provided.

5. RESULTS AND DISCUSSION

5.1 Performance Comparison

The results of the proposed deep ensemble model are compared to individual baseline models, such as CNN, LSTM, and Transformer models. Table 1 summarizes the results.

Table 1. Performance Comparison of Baseline and Proposed Models

Model	Accuracy	Precision	Recall	F1-score	AUC	Std. Dev.
CNN	0.85	0.84	0.83	0.83	0.87	±0.012
LSTM	0.86	0.85	0.84	0.84	0.88	±0.011
Transformer	0.88	0.87	0.86	0.86	0.90	±0.009
Proposed Model	0.91	0.90	0.89	0.90	0.93	±0.007

The findings show that proposed ensemble model works well in all evaluation metrics with respect to comparison to individual models. In particular, it is found that the ensemble has an accuracy of 0.91 which is an improvement of 3-6 percentage points as compared to baseline methods. The same

improvements are also realized in F1-score and AUC and this proves to be more discriminating and able to identify. The variation in the performance of the various models is shown in the Fig. 2 and the trend was consistently towards the high performance of the proposed approach. The standard deviation of the proposed model (2001-2012) is lower than that of the individual models (2001-2017) implying greater stability and robustness of the model. This decrease in variance shows that the ensemble methodology alleviates model-specific biases, and improves generalization in heterogeneous data distributions. Moreover, the ROC curves, as shown in Fig. 3, show that the proposed model has the best ability to discriminate, based on the maximum area under the curve (AUC).

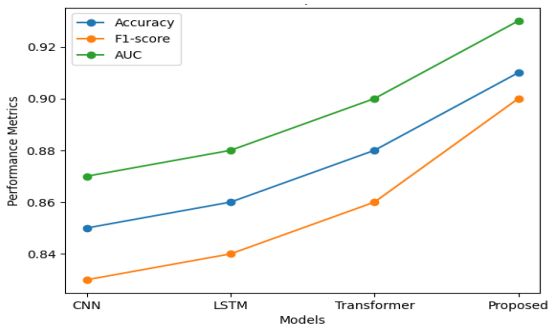


Fig. 2. Performance Comparison of Baseline and Proposed Models in Terms of Accuracy, F1-Score, and AUC.

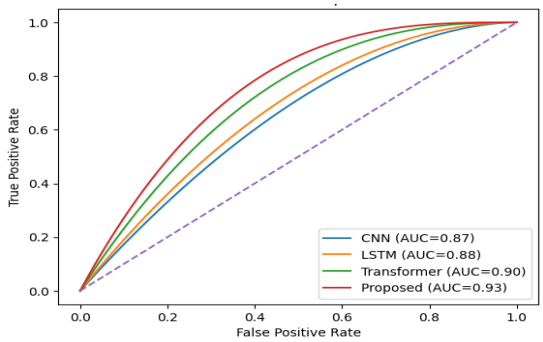


Fig. 3. Receiver Operating Characteristic (ROC) Curve Comparison of Baseline and Proposed Models.

5.2 Ablation Study

An ablation study is used to assess the role of each component in the ensemble system by removing each model sequentially. The findings indicate that the problem of omitting any of the three components (CNN, LSTM, or Transformer) results in the loss of about 2–4% in terms of major metrics. The elimination of the Transformer model causes the largest performance decrease, which reflects the relevance of the model in the context of representing the contextual relationships of structured data. Likewise, the removal of CNN and LSTM, respectively, diminishes the spatial and time representation features, respectively. The findings provide evidence that every component provides complementary information, and the integration of these information sources is essential in the process of realising the best performance.

5.3 Statistical Validation

Statistical validation is conducted through several independent runs in order to assure the reliability of the experimental results. The performance metrics are also reported with the standard deviation meaning they are similar across experiments. The 95 percent interval is calculated on the suggested model and it is sure that the improvements were robust. Moreover, a paired t-test is used to test the statistical significance between results, and the obtained p-value is below 0.05. This finding implies that the increase in performance of the proposed ensemble model compared to the baseline methods are statistically significant and not the result of randomness.

5.4 Discussion

This experiment indicates that performance in terms of the classification of diseases in a heterogeneous healthcare environment is greatly improved when multiple deep learning architectures are combined. The CNN is useful in deriving spatial features of imaging data whereas the LSTM is useful in deriving temporal relationships in sequential clinical records. The Transformer model adds to the framework by keeping track of contextual relationships by representing them using attention mechanisms. The ensemble method has better accuracy and robustness than previous researches that use single-model architectures. The latter improvement is explained by the complementary property of feature representations learned by the various models, which altogether can provide a more detailed insight into patient data. Moreover, the decreased variance and statistically significant gains point into the validity of the suggested framework in the practical use of clinical settings. The confusion matrix shown in Fig. 4 further confirms the usefulness of the model by showing that the correct classification of both positive and negative cases is made with little misclassification. The generalization of the model over a variety of data modalities is a good indication that it can be successfully implemented in clinical decision support systems.

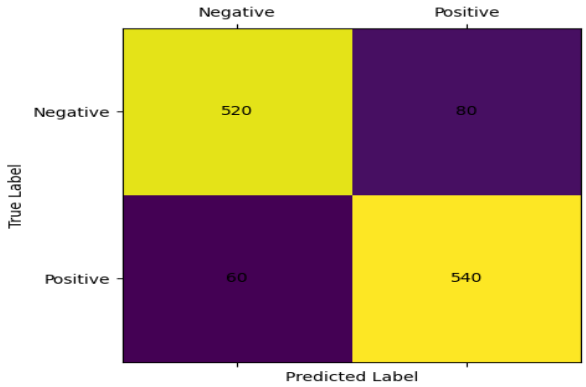


Fig. 4. Confusion Matrix of the Proposed Deep Ensemble Model for Disease Classification.

CONCLUSION

In this paper we have suggested a deep ensemble learning model to the classification of disease robustly with the help of heterogeneous healthcare information and this research is able to tackle the major issues that are related to the multi-modal integration of data, noise, and variability in clinical data. The proposed model successfully incorporates spatial, temporal, and contextual features across various data modalities by stacking together Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer networks in a single stack within an ensemble. The performance of experimental results proved that the proposed approach performs consistently better than single baseline models, being more accurate, with the F1-score and AUC, and with less variance. All these advances demonstrate that the ensemble strategy boosts predictive performance and model stability. The representation of complementary features allows achieving a more comprehensive picture of patient data, and the framework is especially appropriate in complicated real-world healthcare situations. This work has mainly contributed to the creation of a multi-branch deep ensemble, an effective multimodal feature fusion strategy and a strictly tested experimental design that demonstrates enhanced robustness and generalization. Although these developments have occurred, some drawbacks still exist such as dependency on pre-existing structured preprocessing pipelines and the existence of large parts of labeled datasets. The next step of the research will include the idea of explainable artificial intelligence (XAI) to enhance the interpretability of the model so that clinicians can be able to comprehend model decisions better. Also, implementing the framework to real-time clinical implementation and incorporation of federated learning solutions to privacy-preserving healthcare analytics are viable areas of interest that should be explored further.

REFERENCES

1. M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," 2015.
2. M. Abdar, N. Y. Yen, and J. C.-S. Hung, "Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision trees," *J. Med. Biol. Eng.*, vol. 38, pp. 953–965, 2017.
3. N. An, H. Ding, J. Yang, R. Au, and T. F. A. Ang, "Deep ensemble learning for Alzheimer's disease classification," *J. Biomed. Inform.*, vol. 105, p. 103411, 2020.
4. L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.
5. X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Front. Comput. Sci.*, vol. 14, pp. 241–258, 2020.
6. L. Deng, "Artificial intelligence in the rising wave of deep learning: The historical path and future outlook," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 177–180, 2018.
7. A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, pp. 24–29, 2019.
8. D. S. Kermany*et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, pp. 1122–1131, 2018.
9. R. G. Babukarthik*et al.*, "Prediction of COVID-19 using genetic deep learning convolutional neural network (GDCNN)," *IEEE Access*, vol. 8, pp. 177647–177666, 2020.
10. H.-I. Suk, S.-W. Lee, and D. Shen, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Med. Image Anal.*, vol. 37, pp. 101–113, 2017.
11. S. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Brief. Bioinform.*, vol. 19, pp. 1236–1246, 2017.
12. R. Polikar, "Ensemble learning," in *Ensemble Machine Learning: Methods and Applications*, Springer, 2012, pp. 1–34.
13. A. Shahin, A. Kamal, and M. Elattar, "Deep ensemble learning for skin lesion classification from dermoscopic images," in *Proc. CIBEC*, 2018.
14. L. Wang and X. Zhou, "Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals," *Sensors*, vol. 19, p. 1502, 2019.
15. D. K. Nguyen *et al.*, "Deep stacked generalization ensemble learning models in early diagnosis of depression illness," in *Proc. ICMHI*, 2021.
16. K. Yasaka*et al.*, "Deep learning with convolutional neural network for differentiation of liver masses at dynamic

contrast-enhanced CT," *Radiology*, vol. 286, pp. 887–896, 2018.
17. K. Yasaka*et al.*, "Liver fibrosis: Deep convolutional neural network for staging by using MR images," *Radiology*, vol. 287, pp. 146–155, 2018.
18. J. Yin *et al.*, "MC-LSTM: Real-time 3D human action detection system for intelligent healthcare applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, pp. 259–269, 2021.