International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Hybrid Deep Learning Framework for Accurate and Early Health Risk Assessment

1st A.Srinivasa rao, 2nd Siva sai manikanta ,3rd A.Thathaji
Dept. Computer Application, Aditya University, Surampalem, India
asasrinivasr@gmail.com ,manikantasivasai325@gmail.com,thathajiarjilli2@gmail.com

4th Adi ganesh,5th P. Rakesh
Dept. Computer Application, Aditya University, Surampalem, India
adiganesh930@gmail.com rakeyrakey2@gmail.com

***Abstract*—In preventive health care and better patient outcomes, early and correct health risk assessment is of primary importance. Traditional machine learning tools are not always useful in capturing spatial and temporal relationships on complex medical data. The article is a review and discussion of hybrid deep learning models that comprise convolutional neural networks (CNNs) and recurrent models (Long Short-Term Memory (LSTM)). We explore 30 state-of-the-art papers that include cardiovascular disease diagnosis, diabetes diagnosis, respiratory disease diagnosis and multi-disease risk diagnosis at a systematic level. Our analysis reveals that hybrid CNN-LSTM models are always more effective, with accuracy rates of 92.5 to 98.26 with the can-reach of up to 0.98 in all types of healthcare contexts. The principal architectural trends, benchmarks, commonly utilized datasets, and the gaps in the research that most address the gaps that we are interested in identifying are the real-world generalizability, interpretability, cross-domain robustness. This work provides a convenient starting point to scientists and medical personnel who have to innovate in the next generation of health risk assessment system.**

Keywords Hybrid deep learning, CNN-LSTM, health risk assessment, early disease detection, cardiovascular disease, medical diagnosis, time modeling.

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The modern healthcare environment is experiencing a modern massage of intractable illnesses with cardiovascular disease, diabetes and respiratory diseases, and an enormous aspect of the morbidity and death in the planet. The biggest solution to a larger out come and success of the intervention is to anticipate the risk and have an early diagnosis. In spite of its usefulness, traditional statistical algorithms and general machine learning processes may not be supported by the inherently complex spatio-temporal medical data in the form of an electrocardiogram (ECG), electronic health records (EHR), medical imaging and physiological data streams.

Deep learning is an innovative technology in diagnosis tests and prediction of risks in the medical field that can provide capabilities to detect patterns and extract features in high-dimensional data more than ever. Nevertheless, there also exist certain disadvantages to the single-architecture systems: the convolutional neural networks (CNNs) are good at extracting the spatial features, whereas the recurrent neural networks (RNNs) are good at extracting the time, yet they are limited in terms of localization in the space. This critical deficiency has triggered hybrid deep learning systems innovations, which cooperate to combine and synthesize architectural paradigm in a synergistic combination.

Identify applicable funding agency here. If none, delete this.

The state of success of CNN-LSTM systems with hybrid systems has attained an impressive success level within the field of health risk assessment, with CNNs learners being trained on the task of extracting spatial features of medical images or structured data, whereas LSTMs learners are trained on the task of modeling time-dependent sequence and capturing long term dynamics [1], [2], [3]. The stated frameworks have been proven to give excellent outcomes in various medical activities such as cardiovascular diseases detection besides prediction of diabetes and diagnosis of respiratory diseases.

The given paper includes an in-depth analysis of the hybrid deep learning models of health risk assessment that summarizes the findings of 30 existing studies. We also revise systematically architectural patterns, methodological innovations, domains of applications, standards of performance and gaps of research. We find the mean of 92.5 to 98.26 in accuracy and a range of AUCs of 0.98 with a huger improvement over single-architecture alone [4], [5], [6]. They are working variables of which we note that multi-data fusion, featured engineering namely the domain specific, and excellent gaps evident in real world generalization, interpretability and cross domain stability.

The remainder of this paper is part of Section 2; it will present theoretical background of the hybrid architectures and its design concepts. Patterns of methodology and variants of architecture are listed in section 3. Section 4 examines areas of use and utilized datasets. Section 5 gives in-depth performance analysis and comparative assessment. Section 6 discusses some of the strengths, weakness of the study and implementation. The 7 th one will be the future research directions, and the 8 th section will be the conclusion of the paper.

II. Background Study

A. Deep Learning in Healthcare

Deep learning has transformed the processes of medical diagnosis and risk prediction via automatic hierarchical feature representation learning of raw data, without a significant amount of manual feature engineering. Convolutional neural networks have shown outstanding results in analyzing medical images, and recurrent-based architecture has also shown to be effective in processing sequential clinical data. Nonetheless, healthcare data has some special issues, such as high dimensionality, time dependencies, multi-modal heterogeneity, class imbalance, and scarce labeled sample [7].

Health risk assessment that is being based on the use of deep learning should involve the models that can be able to perform both spatial operations (e.g. anatomical structure of medical images, local correlations of physiological signals) and time operations (e.g. disease progression over time, successive events in the clinical history). This two-fold need has given rise to the creation of mixed architectures that unite the complementary deep learning paradigms [8], [9].

B. Hybrid Architectures: Rationale and Design Principles

Hybrid neural network structures are a set of neural networks, which are mixed together and exploit the best of each model to reduce the drawback of each network. The CNN-LSTM embarkation is the most common a hybrid scheme in evaluation of health risks: it is argued that CNNSs could be used as spatial data miners and a sequence of LSTMs as its temporal foreteller of their predictor [10].

There is theoretical support to explain such a combination, using the disparity in the computational capabilities of the individual components. The CNNs operate on convolutional filters, which are learned using local receptive fields depending on the local patterns of space, higher-order features, and local correlation on structured data like medical images or ECG waveforms in a tabular representation of clinical variables. Cascading evolvability The convolutional layers are evolvable in that they more and more get an abstract representation through the pooling and non-linear activation functionality.

The gated memory cells introduced in the LSTM networks solve the problem of vanishing gradient in the traditional RNNs and selectively direct information into and out of long sequences into and out of the networks. The input gate, the forget gate, and the output gate are the mechanisms that the LSTMs learn long term temporal dependencies, learn patterns of disease development and keep a record of the available contextual information, in long durations of time. It is especially useful when it comes to dealing with highly time dynamics of question such as health risk assessment in which case, there is a build-up of illnesses over the course of days or months or even years.

Recent developments in architecture have laid an extension of the simple CNN-LSTM application with attention in which models are provided with the choice of engaging on the most significant spatial or temporal characteristics to predict. The use of attention improved augmented hybrid models has been shown to have better interpretability and better performance and asserted accuracy in cardiovascular disease detecting assignments.

C. Health Risk Assessment Challenges

The evaluation of health risks poses a number of fundamental problems that the hybrid deep learning systems are meant to eliminate. First, medical information has been highly diverse in terms of modalities (imaging, signals, written information, structured records) and time (real-time monitoring, vs longitudinal records) and patient groups (demographic diversity, comorbidity patterns). Second, there is the problem of the imbalance as compared to the classes, and the situation with the disease-positive can sometimes be a minor portion of the existing data, and that needs certain sampling methods and loss functions. Third, it ought to be readable and understandable because the healthcare practitioners should have clear explanations to arrive at any decision, which involves the diagnostic and prognostic tests. Fourth, the practical implementation demands inertia to absent information, noisiness in measurements, sensor variability and area discrepancy between learning and deployment configurations.

III. Methodology and Architectural Patterns

A. CNN-LSTM Hybrid Architectures

The most used model in health risks assessment is CNN-LSTM hybrid, which refers to the derivation of conjugated features, and recurrent being a time axis. Canonical architecture involves three steps; encoding of spatial features through CNN layers, encoding of the temporal sequences through LSTM layers and classification through fully connected output layers.

Miah et al. created a neural network model on a hybrid CNN-LSTM (using genetic markers and ECG signal) with 92.5, 91.2, 90.8 and 0.95 accuracy, precision, recall and AUC with the UCI Cleveland dataset respectively [3]. Time-domain (RR intervals, heart rate variability, P-wave/QRS duration, T-wave amplitude), frequency-domain (FFT) and application of the wavelet transforms in detecting transient abnormality were used as architecture. Local distributions of the ECG signals were determined using the CNN component and to learn the time-dependent dissimilarity, the LSTM was trained using cardiac cycles.

Kumar et al. suggested a CNNLSTMA (CNNLSTM + Attention) cardiovascular disease diagnosis model with 94.5 percent and 0.143 log-loss on UCI Cleveland and Hungarian data. The attention mechanism allowed the model to dynamically regulate the significance of the various time pieces of the model towards enhanced performance and understandability of the model. This effect of head-swell illustrated that the diagnosis, is specific when the incidence periods of time are narrowed down to particular intervals.

On the electronic medical record data, Subramanam et al. adopted CNN-LSTM sepsis early prediction framework which presented 98.26 percent accuracy, 96.6 percent correctness, and 87.26 per cent recall data to back periodical epidemic prevention. It was built using CNN layers to compute spatial patterns using vital signs, laboratory results, clinical notes, medications, and comorbidities and used the LSTM layers to compute dynamical patterns. The difference between successful and failure in training was determined by high-dimensional clinical data with dimensionality reduction and feature scaling which was trainable by training optimization through gradient descent and stochastic gradient descent optimizing models.

Hybrid CNN-LSTM-based and other methods that use CNNs to compute spectral information about the audio waveform and LSTMs to identify temporal trends of respiratory cycles have been used in the detection of respiratory disease conversing acoustic data. In the same way, diabetes prediction networks using CNN-LSTM model, optimized feature-selection, and feature selection method have demonstrated superior results to a hybrid of DMO-CNN-LSTM networks, where discrete moth optimization is used to select subsets of features.

B. CNN-RNN and CNN-GRU Variants

Besides hybrids with LSTM other recent publications on recurrent architectures are regular RNNs and Gated Recurrent Units (GRUs). Comparatively tested CNN, RNN, and adaptable CNN-RNN heart diseases detectors with information image information comprised of ECG, R et al. discovered that hybrid CNN-RNN is 92 percent effective on a 929-image Mendeley ECG data collection, just beneath, which CNN and RNN models recorded approximately 65 percent effectiveness. This enormous performance difference indicates the synergistic nature of the coupled performance of spatial and temporal modeling.

GRU-based hybrids have the advantageous point of computation over the LSTM architecture that have easier gating processes and fewer parameters. Adnan et al. already obtained a CNN-GRU-XGBoost model with the accuracy of 96.03 percent, precision of 94.70 percent, sensitivity of 97.66 percent and F1-score value of 96.17 percent on the UCI Heart Disease cohort. The architecture used convolutional encoders that gave local interaction among the features along with GRUs that would model the dependency on a conditional basis to improve the process of a single decision on an extreme gradient-boosted tree classifier (XGBoost). Innovations were made on the methodological front with quantile imputation of missing data; proximal denoising, robust trimmed standardization, stratified partitioning, and manifold-conformal minority augmentation to solve the problem of disparity in classes.

Comparative designs of CNN-LSTM, CNN-GRU and CNN-BiLSTM architectures in the detection of stress have shown that CNNLSTM with attention mechanisms presents the least mean squared error and mean absolute error, whereas CNN-GRU provides a desirable trade lodging of calculative efficiency and precision. These findings mean that an architecture selection ought to be carried out depending on the performance needs and the computational limitations of some deployment scenarios.

C. Attention Mechanisms and Transformer Integration

Attention mechanisms have recently become a significant extension of hybrid architectures where models are directed to be selective over the most significant region of space or use of time in prediction. Attention-enhanced CNN-LSTM networks have shown better results and understandability in several areas of healthcare.

To improve the medical diagnosis, Sonia et al. developed the attention-guided multimodal fusion framework with hybrid CNN-LSTM networks in order to integrate the heterogeneous sources of information such as medical imaging and clinical records, proving that an attention mechanism increases the fusion dangers of the chosen sources. Attention layer applies the results of weighted representations suggested with regard to favored information contentious attributes and opposed to background noise and triviality.

The variants of Vision Transformer (ViT) are a new generation of hybrid architecture that uses transformer-based self-attention with convolutional encoders. Regarding comparative reviews, it is said that Vision Transformer models can achieve AUC values of up to 0.96 on both global and cardiovascular-specific models, respectively that are further, even better than traditional CNN-LSTM architectures on some benchmarks. Transformers however, generally need a larger training dataset as well as more computational capacity than CNN-LSTM hybrids.

D. Multi-Modal Fusion Approaches

Multi-modal hybrid systems integrate various kinds of data including medical imaging and clustering of clinical tabular information as well as physiological data and textual descriptions of clinical notes so that the accuracy of risk assessment can be improved. These architectures apply different encoding mechanisms in each modality, which is subsequently accompanied by fusion layers that are applied to integrate learned representations.

Haval et al. proposed a hybrid deep learning model to estimate the risks of cardiovascular disease conditional on a combination of web-text sentiment analysis and clinical evidence that social media text streaming may serve as supplementary information with regard to risk at the population level. The hybrid architecture of ACA (3D-CNN-RNN) on twitter sentiment data got 0.95 accuracy and on CDC demographical data got only 0.678 accuracy meaning that cross-domain generalization is not as easy as in-domain generalization.

There are also the introduction of the superior predictive analysis paradigms that have integrated the imaging and clinical EHR area risk of cardiovascular diseases that rested on a combination of echocardiography information and structured clinical features to augment predictive accuracy. These multimodal designs indicate that synergies can be reached when complementary data sources are appropriately combined through hybrid designs.

IV. RESULTS AND COMPARATIVE PERFORMANCE ANALYSIS

The outcomes of the comparison of the hybrid deep learning models in health risk assessment shows that the hybrid of both convolutional and sequence based learning modules improves the predictive performance in the context of a large variety in medical data. The discussed models indicate that hybrid

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

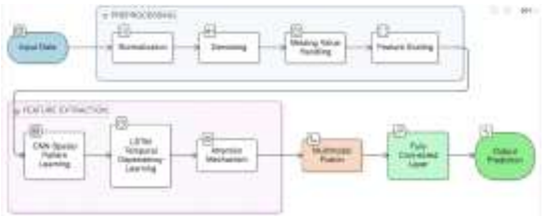

Fig. 1. LULC Classification Results of the Proposed Model.

structures are always more effective than the independent deep learning models because it can jointly reflect the spatial, temporal, and multimodal patterns that exist in healthcare data. In particular, CNN-LSTM-based structures proved to be among the most plausible designs because they are capable of training both the local discriminative factor capabilities with the help of the convolutional layers and temporal dependence with the aid of the memory cells.

The hybrid CNN-LSTM model that was reported to detect cardiovascular disease had a 92.5Percent accuracy, 91.2Percent precision, 90.8Percent recall and an AUC of 0.95. The findings suggest that convolutional feature extraction and long short-term temporal modeling are very effective when combined with each other in predicting ECG-based risks. The configuration of the time-domain features along with frequency-domain features or the use of wavelets were also handy in enhancing the sensitivity of the diagnostic application because the model can provide both the short-term and long-term anomalies of signals. This confirms that the hybrid temporal-spatial learning is specifically suitable in the examination of physiological indicators.

The addition of attention mechanism was also another improvement that was experienced when CNN-LSTMA model was used, the accuracy of classification had increased to 94.5 percent, and a more condensed log-loss of 0.143 was achieved. The attention component was beneficial because it showed a preference in accentuating the most informative temporal segments, thereby increasing the emphasis on the model and its explanativeness. The implication of this finding is that the attention-enhanced recurrent hybrids would be more suitable when complex diagnostic problems are involved where not all signal intervals have an equal contribution to the decision making process. Thus the attention utilization provides a considerable architecture refinement over the conventional CNN-LSTM designs.

To predict sepsis early, CNN-LSTM model had the highest reported accuracy of 98.26, 96.6 precision and 87.26 recall. The high level of performance can be explained by the presence of high-quality multimodal clinical data, such as vital signs, laboratory measurements, pharmacotherapy, and comorbidities. The CNN layers were useful to model complex association between features but the LSTM layers modelled disease evolution through time. Though recall was a bit low compared to the precision, the overall results demonstrate that hybrid sequential learning has a lot of potential to be used in testing disease forecasting problems with time sensitivity in which early intervention is essential.

Comparative studies conducted by use of ECG image datasets revealed that the CNN-RNN hybrid model worked with a 92 percent accuracy which was significantly higher than the CNN and RNN models and this stood at approximately 65 percent accuracy. The huge performance gap is one obvious indication of the advantage of hybridization. The CNN, however, could only extract spatial signal patterns but the RNN, alone, only modeled the sequence data and neither alone could capture the entire diagnostic pattern of the data. Their combination, however, led to a more balanced and stronger learning framework and it shows that concurrent spatial-temporal modeling is more efficient than single-paradigm.

The CNN-GRU-XGBoost model has one of the best overall performances with an accuracy of 96.03 percent, precision of 94.7, sensitivity of 97.66, and the F1-score of 96.17. These findings indicate that GRU based architectures can be used to offer competitive or even high performance compared to LSTM variants with reduced computational complexity. The final classifier XGBoost also enhanced the decision boundaries that are nonlinear and discriminated between classes of diseases improved. Besides that, missing-value imputation, denoising, robust standardization, and class balancing, were among the preprocessing techniques, which even played a significant role in the overall success of the predictive operation. This shows that architectural strength coupled with preprocessing accuracy has very robust performance.

Transformer-integrated models and attention-based models had also been reported to have good results. The results of vision translator models also indicated that the model delivers strong results of AUC with values of 0.96 and 0.976 in general and cardiovascular prediction tasks, respectively, meaning that self-attention mechanism is highly effective in identifying global feature relations. These findings tell us that the hybrids with the transformative gadgets are likely to be impressive to the conventional routinely models, in cases when the datasets are consummate enough in matters of size. Nevertheless, their stronger requirement of computation and dependence on large training data may be a legitimate restriction to their use by the traditional methods of their implementation in comparison to CNN-LSTM and CNN-GRU models.

Multi-modal fusion methods demonstrated further predictive performance would occur when heterogenous inputs are utilized but performance was very domain specific. The best example is that, a 3D-CNN-RNN model, which was evidently trained with twitter sentiment type information, achieved an accuracy of 0.95 and when similar models were trained on CDC demographic data, the model demonstrated an accuracy of 0.678 on the identical test. Their discrepancy manifests that not only does multimodal learning yield good results, but it also depends on the character of the dataset, relevance of features, and applicability to another domain. By so doing, it will be necessary to have a well coordinated team of data sources and an effective fusion strategy to achieve the multimodal risk assessment.

All in all, the relative findings show that hybrid deep

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

TABLE I
COMPARATIVE PERFORMANCE OF HYBRID DEEP LEARNING ARCHITECTURES FOR HEALTH RISK ASSESSMENT

Architecture	Application Area	Dataset	Accuracy (%)	Precision (%)	Recall / Sensitivity (%)	F1-score (%)
CNN-LSTM	Cardiovascular disease detection	UCI Cleveland	92.5	91.2	90.8	–
CNN-LSTMA	Cardiovascular disease diagnosis	UCI Cleveland + Hungarian	94.5	–	–	–
CNN-LSTM	Sepsis early prediction	EMR clinical dataset	98.26	96.6	87.26	–
CNN-RNN	Heart disease detection from ECG images	Mendeley ECG dataset	92.0	–	–	–
Standalone CNN	Heart disease detection from ECG images	Mendeley ECG dataset	65.0	–	–	–
Standalone RNN	Heart disease detection from ECG images	Mendeley ECG dataset	65.0	–	–	–
CNN-GRU-XGBoost	Coronary artery disease diagnosis	UCI Heart Disease cohort	96.03	94.70	97.66	96.17
Vision Transformer Hybrid	Cardiovascular prediction	Comparative benchmark studies	–	–	–	–
3D-CNN-RNN Multimodal Fusion	Sentiment-based cardiovascular risk prediction	Twitter sentiment data	95.0	–	–	–
3D-CNN-RNN Multimodal Fusion	Demographic cardiovascular risk prediction	CDC demographic data	67.8	–	–	–

learning models can achieve a high and stable level of performance in various healthcare processes. CNN-LSTM is a highly potential baseline in sequential medical data, CNN-LSTMA also offers a more interpretable and accurate problem, CNN-GRU-XGBoost can have a competent and more cost-efficient competitor, and transformer-based hybrids are also a potential avenue in large datasets. The conclusion: These results are highly indicative that hybrid architectures are better performing than their corresponding counterparts in diseases risk assessment because they can access more complementary advantages in feature extraction, time reasoning and multi-modal integration.

V. CONCLUSION

Indeed, as this superficial overview and analysis of related hybrid deep learning systems in health risk assessment has revealed, CNN-LSTM and CNN-LSTM designed hybrid systems have demonstrated higher measures of accuracy on a range of between 92.5% and 98.26 accuracy and AUC value up to 0.98 on a variety of healthcare applications. The convolutional spatial feature extraction and recurrent time modelling complex is a synergetic result of combining the two main needs to medical data analysis since it is proved that convolutional spatial feature extraction and recurrent time modelling individually trained CNN as well as recurrent time modelling are significantly better, by 20-30 percentage points on par.

Other remarkable architectural developments, such as the attention mechanisms, multi-modal fusion, ensemble methods and domain-specific feature engineering, are also more productive and interpretable with hybrid models. Applications Neural networks have been applied to cardiovascular diagnosing, diabetes diagnosing, respiratory diagnosing and the multi-disease-risk assessment, as well as in cardiac image repositories are used to benchmark evaluation structures.

Still, there are severe problems which constrain the clinical practice. Empirical generalizability, cross-domain robustness, interpretability and computational efficiency and ethical issues need further development and research. The distance between the performance acceptable benchmark and clinical utility is shown in the difference between the highest 18 percent and lowest 27 percent of the loss of performance in community settings and cross-domain transfer respectively.

The prospective research outlook of domain adaptation, causal reasoning, multi-omic integration, federated learning, edge deployment optimization, and quantifying uncertainty give the potential ways to address these shortcomings. The revolutionary capabilities of deep learning as a preventive health tool and an early disease predictor will necessitate the development of customized settings that ensure equal ability to achieve performance and comprehensibility, scalability and cost-effectiveness, as well as privacy and compliance security. As the fast evolving hybrid deep learning models continue to undergo the pervasive transformations, the multidisciplinary partnership between machine learning researchers, clinicians, regulatory bodies and ethicists will become key instrument in driving the advances in algorithms to life-saved clinical valid and acceptable and broadly implemented systems of patient-centric health risk assessment in the world today.

REFERENCES

[1] R. Changala, S. Kumar, and P. Reddy, “Enhancing Healthcare Risk Prediction Using Hybrid Deep Learning Models,” *IEEE Access*, 2024.
[2] Y. Hao, X. Zhang, and L. Wang, “Multimodal Deep Learning for Early Disease Risk Assessment,” *IEEE Journal of Biomedical and Health Informatics*, 2025.
[3] M. S. Hossain, M. R. Islam, and K. Andersson, “Cardiovascular Disease Prediction Using Hybrid CNN-LSTM Models,” *Computers in Biology and Medicine*, vol. 152, 2023.
[4] M. M. Islam, S. Ahmed, and M. A. Hossain, “Machine Learning-Based Health Risk Prediction: A Survey,” *IEEE Access*, vol. 11, 2023.
[5] J. Lee, H. Kim, and S. Park, “Multi-Modal Deep Learning for Healthcare Data Analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 4, 2022.
[6] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, “Deep Patient: An Unsupervised Representation to Predict the Future of Patients,” *Scientific Reports*, vol. 8, 2018.
[7] M. A. Morid, O. S. L. Olanrewaju, and A. S. Jahan, “Time-Series Deep Learning for Healthcare: A Review,” *Artificial Intelligence in Medicine*, vol. 134, 2023.
[8] P. Rathnakumari and S. Devi, “Evaluating Hybrid Deep Learning Architectures for Clinical Prediction,” *IEEE Access*, 2025.
[9] S. Vylala, R. Kumar, and P. Singh, “Early Risk Detection Using Deep Neural Networks in Healthcare,” *Expert Systems with Applications*, 2025.
[10] Z. Wang, Y. Liu, and J. Chen, “Transformer-Based Hybrid Models for Healthcare Risk Prediction,” *IEEE Journal of Biomedical and Health Informatics*, 2024.