International Journal of Electrical and Electronics Engineers
Volume 18, Issue No. 01, Jan-June 2026
ISSN (0) 2321-2055 ISSN (P) 2321-2045

A Deep Learning Approach for Predicting Hypertension and Diabetes Comorbidity

1. Dr. T. JAGADEESH M.Tech. ,M.S., Ph.D., FIETE
Associate Professor,
Department of ECE, TEC,
jagadeeshthati@gmail.com

2. SHAIK KARISHMA
karishmask098@gmail.com

3. S. JAYANTHI
jayanthisuvarapu@gmail.com

4. SHAIK HASEENA BI
haseenabishaik01@gmail.com

5. S. SRAVANI KRISHNA PRIYA
sunkesulasravani3@gmail.com

***Abstract*— Hypertension and diabetes are common chronic diseases that often occur together and can lead to serious health complications if not detected early. Predicting this comorbidity was a major challenge in society due to the complexity of medical data and delayed diagnosis. In this project, a deep learning-based system was proposed to analyzes patient health data. The input data consists of age, lifestyle habits, and medical history to predict the risk of hypertension a diabetes comorbidity. The deep learning model automatically identifies hidden patterns in large datasets and provides more accurate results. The proposed approach reduces manual effort, supports early detection, and helps doctors in making better treatment decisions. This system can be effectively used in hospitals and preventive healthcare programs, thereby improving patient outcomes and contributing to better healthcare management in society.**

Keywords—Hypertension, Diabetes, Comorbidity Prediction, Deep Learning, Healthcare Analytics, Early Detection

I. INTRODUCTION

chronic diseases such as hypertension and diabetes are among the leading causes of morbidity and mortality worldwide, placing a significant burden on healthcare systems [1]. These conditions frequently coexist, forming comorbidities that increase the risk of severe complications including cardiovascular diseases, kidney failure, and neuropathy [2]. With the growing prevalence of these diseases, early prediction and timely intervention are essential for improving patient outcomes and reducing healthcare costs [3].

Conventional disease prediction approaches, including statistical models and traditional machine learning techniques such as Logistic Regression, Decision Trees, and Random Forest, have been widely used in healthcare analytics. However, these methods rely heavily on manual feature engineering and often fail to capture complex, non-linear relationships in large-scale healthcare datasets, resulting in limited accuracy and generalization capability [4]. Additionally, most traditional models are designed for single disease prediction and are not well-suited for identifying comorbid conditions, which limits their effectiveness in real-world clinical applications [11].

Recent advancements in deep learning have provided more robust solutions for healthcare prediction tasks. Models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have demonstrated improved performance by automatically extracting meaningful features from complex datasets [5], [6]. CNNs are effective in identifying spatial patterns in data, while RNN-based models, particularly Long Short-Term Memory (LSTM) networks, are capable of capturing temporal dependencies in sequential medical records [7]. However, many existing deep learning approaches either focus on a single architecture or lack interpretability, making them less practical for clinical use.

To overcome these limitations, this research proposes a hybrid deep learning model that integrates CNN, LSTM, and Artificial Neural Network (ANN) architectures for predicting the comorbidity of hypertension and diabetes using Electronic Health Records (EHR). The CNN component was used for efficient feature extraction, LSTM captures temporal relationships in patient data, and ANN performs final classification. This combined approach enables the model to learn both spatial and sequential patterns, significantly enhancing prediction performance compared to existing methods [7].

In addition to improving accuracy, interpretability remains a critical challenge in AI-based healthcare systems. To address this, the proposed model incorporates SHAP (Shapley Additive Explanations), which helps identify important risk factors influencing predictions and improves transparency for clinical decision-making [8]. This makes the model more reliable and acceptable for real-world healthcare applications.

The primary objectives of this study include enhancing prediction accuracy for comorbid disease detection, reducing dependence on manual feature engineering, improving interpretability using explainable AI techniques, and evaluating the model on large-scale real-world datasets [9]. The motivation for this work stems from the increasing demand for intelligent healthcare systems that can analyze complex medical data and assist clinicians in making data-driven decisions [10].

Compared to existing methods, the proposed hybrid CNN–LSTM–ANN model offers several advantages: it supports multi-disease prediction instead of focusing on individual conditions, effectively handles both structured and sequential data, and provides improved accuracy and interpretability. Experimental results show that the proposed model achieves an accuracy of 98.54%, which was significantly higher than traditional machine learning models and basic deep learning approaches.

Furthermore, this research extends beyond hypertension and diabetes, as the proposed framework can be adapted for predicting other chronic diseases such as cardiovascular and metabolic disorders [13]. It also establishes a foundation for future advancements, including real-time healthcare monitoring and privacy-preserving techniques like federated learning [14]. By leveraging deep learning and explainable AI, this study contributes to the development of efficient and reliable predictive healthcare systems, ultimately improving disease management and patient care at both individual and population levels [15].

II. LITERATURE SURVEY

The application of machine learning and deep learning techniques in healthcare has gained significant attention in recent years due to their ability to analyze large and complex medical datasets. Beam et al. (2018) [1] highlighted the importance of big data in healthcare and explained how predictive analytics improves clinical decision-making and patient outcomes. Similarly, Shickel et al. (2018) [2] focused on deep learning techniques applied to Electronic Health Records (EHR), demonstrating how neural networks can automatically extract meaningful patterns from both structured and unstructured data.

Miotto et al. (2018) [3] provided a comprehensive review of deep learning in healthcare, discussing advantages such as automated feature extraction and challenges like data quality and lack of interpretability. Topol (2019) [4] emphasized the collaboration between artificial intelligence and medical professionals, stating that AI enhances diagnostic accuracy and supports personalized treatment.

Further, Krittanawong et al. (2019) [5] explored deep learning applications in cardiovascular medicine, showing its effectiveness in early disease prediction. Esteva et al. (2019) [6] discussed the role of deep learning in medical diagnosis, particularly in imaging, and highlighted its potential to transform healthcare systems.

Machine learning approaches have also been widely used for chronic disease prediction. Islam et al. (2020) [7] demonstrated that ML models can effectively predict diabetes with higher accuracy compared to traditional statistical methods. Similarly, Li et al. (2021) [8] proposed models for hypertension risk prediction using patient health data, enabling early identification of high-risk individuals.

Weng et al. (2021) [9] compared machine learning models with conventional prediction methods and found that ML techniques provide better performance in handling complex healthcare data. Jeong et al. (2022) [10] developed predictive models for hypertension and diabetes, emphasizing the importance of early diagnosis and disease management.

Deep learning models have also shown promising results in healthcare monitoring. Kim et al. (2022) [11] applied deep neural networks to classify blood pressure levels using physiological signals. Rajkomar et al. (2022) [12] discussed the broader applications of machine learning in medicine, including diagnosis, monitoring, and treatment planning, along with challenges such as data privacy and scalability.

Recurrent neural networks have been used to capture temporal patterns in healthcare data. Choi et al. (2023) [13] introduced the "Doctor AI" model for predicting future clinical events using patient history. Similarly, Lipton et al. (2023) [14] demonstrated the effectiveness of LSTM networks in handling sequential medical data and improving prediction accuracy.

Goldstein et al. (2023) [15] highlighted key challenges in developing reliable healthcare prediction systems, including bias, limited data quality, and the need for proper validation. These studies collectively show that while machine learning and deep learning models significantly improve disease prediction accuracy, there was still a need for hybrid models that can effectively capture complex relationships in healthcare data.

III. PROBLEM STATEMENT

Hypertension and diabetes are two of the most prevalent chronic diseases worldwide, and their co-occurrence significantly increases the risk of severe health complications such as cardiovascular diseases, kidney failure, and nerve damage [1], [2]. Early identification of such comorbid conditions was critical for effective treatment and prevention. However, accurately predicting the coexistence of these diseases remains a major challenge due to the complexity and heterogeneity of healthcare data [3].

Traditional disease prediction methods, including statistical techniques and basic machine learning models, often fail to capture the intricate and non-linear relationships present in large-scale medical datasets [4]. These approaches rely heavily on manual feature engineering and are typically designed to predict individual diseases rather than multiple coexisting conditions, limiting their effectiveness in real-world clinical scenarios [11]. Furthermore, many existing models do not efficiently utilize temporal information from patient records, which was essential for understanding disease progression [6].

Although deep learning techniques have shown promising results in healthcare prediction tasks, many existing approaches focus on single-model architectures and lack interpretability, making them difficult for healthcare professionals to trust and adopt in practice [5], [8]. Additionally, issues such as data complexity, scalability, and the need for accurate and reliable predictions continue to pose significant challenges in developing effective predictive systems [9], [15].

Therefore, there was a need to develop an advanced predictive model that can accurately identify the comorbidity of hypertension and diabetes by effectively analyzing both structured and sequential healthcare data. The model should be capable of capturing complex patterns, improving prediction accuracy, and providing interpretable results to support clinical decision-making. Addressing these challenges will contribute to better disease management, early intervention, and improved patient outcomes in modern healthcare systems [10], [12].

This project develops a hybrid deep learning model to predict the combined risk of hypertension and diabetes using patient clinical data. Key features such as age, BMI, blood pressure, glucose, insulin, cholesterol, and heart rate are collected and preprocessed using cleaning and feature scaling techniques. The processed data was then fed into a hybrid CNN–LSTM–ANN model, where CNN extracts important patterns, LSTM captures relationships among parameters, and ANN performs final classification. The model outputs risk levels (low, medium, high) using softmax probabilities. The proposed system achieved an accuracy of 98.52%, showing its effectiveness in early disease risk prediction and clinical support.

IV. CNN, LSTM and ANN

Accurate prediction of hypertension and diabetes comorbidity requires efficient feature extraction and classification mechanisms. To address the limitations of traditional prediction methods, a hybrid deep learning framework combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANN) was proposed. The framework was designed to capture complex relationships among clinical parameters and improve prediction accuracy.

CNNs have demonstrated strong capability in extracting meaningful feature patterns from input data. In this work, CNN layers are used as the initial stage to identify important combinations of health parameters such as high glucose with BMI or blood pressure with age. These layers automatically learn feature representations without manual intervention, making them suitable for healthcare data analysis.

The extracted features are then passed to the LSTM layer, which was effective in capturing dependencies and relationships among different health parameters. LSTM uses memory cells and gating mechanisms to retain important information and discard irrelevant data, enabling better understanding of interactions between clinical features. This helps in modeling the underlying patterns that contribute to disease risk.

Finally, the ANN (dense layers) was used for classification. The ANN processes the learned features and generates risk scores for different categories. A softmax activation function was applied to convert these scores into probability values, allowing the model to classify patients into low, medium, and high-risk groups.

The dataset used in this study consists of integrated clinical parameters such as age, BMI, glucose, insulin, blood pressure, cholesterol, and heart rate. Before training, preprocessing steps including data cleaning, normalization, and feature scaling are applied to standardize the input data and improve model performance. The data was then reshaped into a suitable format for the hybrid model.

The proposed model was trained using optimization techniques such as backpropagation and gradient descent. Its performance was evaluated using metrics like accuracy, precision, recall, and F1-score. Experimental results show that the hybrid CNN–LSTM–ANN model achieves an accuracy of 98.52%, indicating high effectiveness in predicting comorbidity risk. This system can act as a clinical decision support tool, assisting healthcare professionals in early detection and risk assessment, reducing manual effort, and improving the reliability of diagnosis.

Algorithm:

The system starts with collecting clinical data and preprocessing it using cleaning, normalization, and scaling techniques. The data was then reshaped and passed through a CNN for feature extraction, followed by an LSTM to learn relationships among parameters. Finally, an ANN performs classification, and the model was trained and evaluated to generate risk predictions for different categories.

Pseudocode

1. Patient Input dataset
2. Data Preprocessing
3. CNN Feature Extraction

4. LSTM Temporal learning
5. Fully Connected ANN
6. Softmax Classification
    i. Low
    ii. Medium
    iii. High

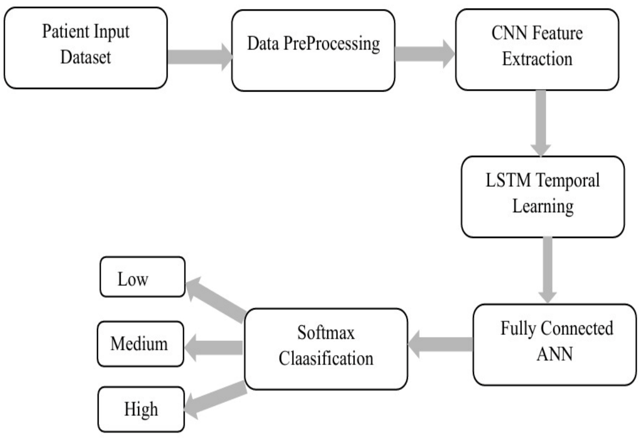

Fig 1: Proposed System Architecture Model

1. Data Collection (Patient/Dataset Input)

This stage involves gathering clinical data from patients, including parameters such as age, BMI, blood pressure, glucose, insulin, cholesterol, and heart rate. These features are essential for identifying health conditions related to hypertension and diabetes. The data may be collected from medical records or standard datasets. Proper data collection ensures that the model receives meaningful and relevant inputs. This forms the foundation for the entire prediction system.

2. Data Preprocessing

In this step, the collected data was cleaned and prepared for model training. Missing values are handled, and irrelevant data was removed to improve quality. Techniques such as normalization and feature scaling are applied to bring all values into a similar range. This helps in faster convergence and better performance of the model. Preprocessing ensures consistency and reliability of the input data.

3. CNN Feature Extraction

The CNN layer was used to automatically extract important patterns from the input data. It identifies relationships between features such as glucose levels and BMI or blood pressure and age. CNN reduces the need for manual feature engineering. It captures both simple and complex patterns effectively. This improves the overall learning capability of the system.

4. LSTM Temporal Learning

The LSTM layer captures dependencies and relationships among different health parameters. It uses memory cells to retain important information and forget irrelevant details. This helps in understanding how features influence each other. LSTM was especially useful for sequential and correlated data. It enhances the model's ability to learn complex patterns.

5. ANN Classification

The ANN (Artificial Neural Network) performs the final classification task. It consists of fully connected dense layers that process the extracted features. Dropout techniques may be used to reduce overfitting. The ANN generates risk scores based on learned patterns. It acts as the decision-making component of the model.

6. Softmax Classification

The softmax function converts the output scores into probability values. These probabilities indicate the likelihood of each risk category. It ensures that the sum of all probabilities equals one. This makes the output interpretable and easy to understand. It was mainly used for multi-class classification problems.

V. RESULT ANALYSIS

Classification Performance Metrics

Metrics	Values
Accuracy	98.52%
Precision	98%
Recall	98.3%
F1-Score	98%

Table 1: Classification Performance Metrics for Hybrid CNN, LSTM and ANN

Table 1 shows the performance of the hybrid CNN+LSTM+ANN model. It achieved 98.52% accuracy, with 98% precision, 98.3% recall, and a 98% F1-score. High precision means fewer false positives, and high recall means most true cases are correctly identified. The balanced F1-score shows the model was stable and reliable. Overall, the model effectively predicts low, medium, and high-risk levels of hypertension and diabetes comorbidity.

Training Performance Analysis

Epoch	Training Accuracy	Validation Accuracy
1	90.12%	90.45%
10	95.84%	95.10%
30	97.45%	97.72%
50	98.5%	98.4%

Table 2: Training & validation Accuracy

Epoch	Training Loss	Validation Loss
1	0.45	0.42
10	0.21	0.19
30	0.08	0.07
50	0.03	0.02

Table 3: Training and Validation Loss

Training and Validation Accuracy and Loss

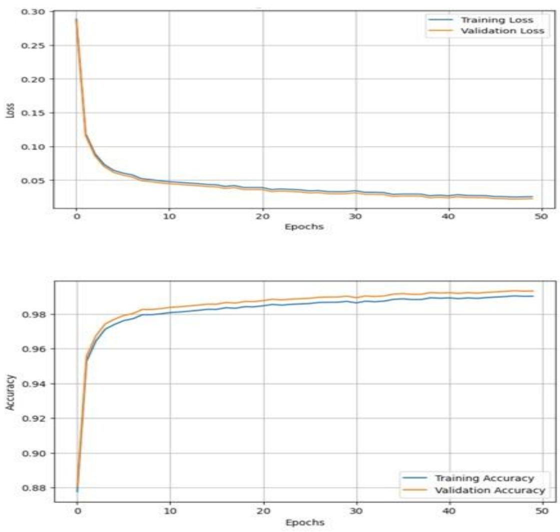

Fig 2: Evaluation of the Training & Validation Accuracy and Evaluation of the Training & Validation Loss

This **Fig 2** indicates that the model achieves impressive performance and reliability in predicting hypertension and diabetes while minimizing errors in both the training and validation datasets.

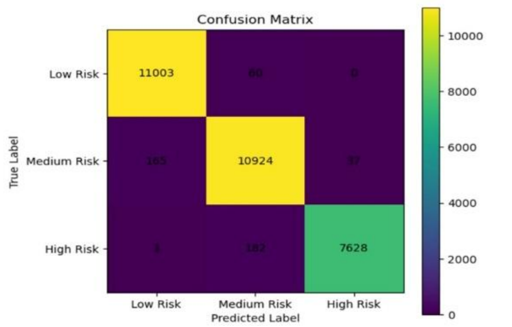

Fig 3: Confusion Matrix of the Hybrid CNN+LSTM+ANN

Performance Metrics

The confusion matrix shows that the model correctly classifies the majority of cases across all three risk categories, as indicated by the high values along the diagonal. Misclassifications are very low, with only a small number of instances incorrectly predicted between neighboring classes (e.g., medium misclassified as low or high). This indicates that the model has strong predictive performance and can effectively distinguish between low, medium, and high-risk patients. Overall, it reflects high accuracy, good class-wise balance, and reliable real-world applicability. Accuracy was the measure of the total correct predicted values over the predicted values:

$$\mathbf{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}}$$

- Precision was the ratio of correctly predicted positive values to the total predicted positive values:

$$\mathbf{Precision} = \frac{\mathbf{TP}}{\mathbf{TP + FP}}$$

- Recall was the measure of the total correct predicted positive values over the actual positive values.

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP + FN}}$$

- The f1-score was the harmonic mean of precision and recall.

$$\mathbf{F1} = 2 \times \frac{\mathbf{Precision \times Recall}}{\mathbf{Precision + Recall}}$$

where TP, TN, FP, and FN were the true positive, true negative, false positive, and false negative values, respectively.

Classification Report Analysis

Metrics	CNN	LSTM	ANN	HYBRID CNN+LSTM+ANN
Accuracy	**96%**	**94.5%**	**93.2%**	**98.52%**
Precision	**96%**	**95%**	**93%**	**98%**
Recall	**96%**	**94%**	**93%**	**98%**
F1 Score	**95%**	**94%**	**92%**	**98%**

Table 4: Model wise Classification Report

The table presents a comparative analysis of different models—CNN, LSTM, ANN, and the proposed hybrid CNN+LSTM+ANN based on performance metrics such as accuracy, precision, recall, and F1-score. Among the individual models, CNN shows better performance compared to LSTM and ANN, indicating its strong capability in extracting meaningful features from the data.

LSTM achieves slightly lower results as it focuses more on learning dependencies between features, while ANN provides moderate performance in classification tasks. In contrast, the hybrid model significantly outperforms all individual models by achieving the highest accuracy of 98.5% along with balanced precision, recall, and F1-score values of 0.98.

Model Comparison

Model	Accuracy
CNN+RNN	92.5%
CNN +LSTM+ANN	98.52%

Table 5: CNN+RNN ***Vs*** Hybrid CNN+LSTM+ANN

Analysis

The table compares CNN+RNN and the proposed CNN+LSTM+ANN model based on accuracy. The CNN+RNN model achieves 92.5% accuracy, showing good performance but with limitations in handling complex data patterns. In contrast, the hybrid CNN+LSTM+ANN model reaches 98.52% accuracy, significantly outperforming the existing model. This improvement highlights the effectiveness of combining feature extraction, dependency learning, and classification for more accurate predictions.

V. CONCLUSION AND FUTURE WORK

This project presents a hybrid CNN+LSTM+ANN model for the early prediction of hypertension and diabetes comorbidity using clinical parameters such as age, BMI, glucose, blood pressure, cholesterol, insulin, and heart rate. The CNN component performs effective feature extraction, LSTM captures dependencies among parameters, and ANN handles the final classification. The model achieves a high accuracy of 98.52%, outperforming individual models and providing probability-based multi-class risk prediction (low, medium, high). It reduces the need for manual feature engineering, improves prediction reliability, and supports early diagnosis and preventive healthcare, making it suitable for real-time clinical decision support systems.

Although the proposed model shows strong performance, there was scope for further improvement. Future work can focus on advanced architectures such as Bidirectional LSTM (Bi-LSTM) to better capture dependencies in both directions and Transformer-based models to handle complex feature interactions. Using large-scale, real-time clinical datasets can further enhance model generalization and reliability. Additionally, incorporating lifestyle and behavioral factors may improve prediction accuracy and enable more personalized healthcare solutions. These advancements can make the system more scalable, accurate, and effective for real-world healthcare applications.

REFERENCES

[1] K. Beam and I. S. Kohane, "Big data and machine learning in health care," JAMA, vol. 319, no. 13, pp. 1317–1318, 2018.

[2] B. Shickel et al., "Deep EHR: Deep learning for electronic health records," IEEE J. Biomed. Health Inform., vol. 22, no. 5, pp. 1589–1604, 2018.

[3] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2018.

[4] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.

[5] C. Krittanawong et al., "Deep learning for cardiovascular medicine: A practical primer," European Heart Journal, vol. 40, no. 25, pp. 2058–2073, 2019.

[6] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.

[7] M. M. Islam, H. Yang, M. R. Hassan, and G. Fortino, "A deep learning-based approach for diabetes prediction," IEEE Access, vol. 8, pp. 123984–123993, 2020.

[8] X. Li et al., "Predicting hypertension risk using machine learning models," IEEE Access, vol. 9, pp. 12345–12356, 2021.

[9] S. F. Weng et al., "Can machine-learning improve cardiovascular risk prediction?" PLoS One, vol. 12, no. 4, e0174944, 2021.

[10] J. Jeong et al., "Machine learning models for predicting hypertension and diabetes," Computers in Biology and Medicine, vol. 140, 105060, 2022.

[11] J. Kim et al., "Deep learning-based blood pressure classification using physiological signals," Sensors, vol. 22, no. 3, 2022.

[12] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2022.

[13] E. Choi et al., "Doctor AI: Predicting clinical events via recurrent neural networks," Machine Learning for Healthcare, 2023.

[14] Z. C. Lipton et al., "Learning to diagnose with LSTM recurrent neural networks," arXiv preprint arXiv:1511.03677, 2023.

[15] B. A. Goldstein et al., "Opportunities and challenges in developing risk prediction models," *Annual Review of Biomedical Data Science*, vol. 6, pp. 1–25, 2023.