International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 4, 2025

An Innovative Ensemble Deep Learning Clinical Decision Support System For Diabetes Prediction

Nada Fayaz [1], Dr. Ijteba Sultana[2]

[1] PG Scholar, [2] Associate Professor
[1,2] Department of Computer Science and Engineering, ISL Engineering College, Hyderabad, Telangana, India
nadafayaz01@gmail.com, drijtebasultana@gmail.com

Abstract: *Diabetes is a significant global health concern, with an increasing number of diabetic individuals at risk. Early prediction of diabetes is essential for preventing its progression and reducing the risk of severe complications such as kidney and heart diseases. This study proposes an innovative Ensemble Deep Learning (EDL) clinical decision support system for diabetes prediction with high accuracy. Three diabetes datasets are utilized: the Pima Indian Diabetes Dataset, the Diabetes Dataset from Frankfurt Hospital, Germany, and the Iraqi Diabetes Patient Dataset (IDP, to train the novel EDL models. The Extra Tree Classifier (ETC) approach is employed to extract relevant features from the data. The performance of the proposed EDL models is rigorously evaluated based on key metrics such as accuracy, precision, recall, and F-score. The study employs a variety of algorithms, including ANN, CNN, LSTM, Stack-LSTM, Stack-ANN, Stack-CNN, and CNN + LSTM, with further potential to enhance performance by exploring additional ensemble techniques such as CNN + LSTM.*

Keywords*: Ensemble Deep Learning (EDL), ANN, CNN, LSTM.*

I. INTRODUCTION

Diabetes Mellitus (DM) is a metabolic disorder where the body is unable to use glucose appropriately and as a result, the level of glucose in the blood remains high. Complications such as diabetic ketoacidosis, chronic renal failure, nonketotic hyperosmolar coma, foot ulcers, eye damage, heart disease, stroke and chronic renal failure are some of the serious problems that diabetes can bring. The three most common DM include Type 1 Diabetes (T1D), Type 2 Diabetes (T2D) and Gestational Diabetes. T1D occurs as a result of the body failing to produce sufficient insulin. The increasing number of diabetic patients is putting them in danger and this is why diabetes is a health issue of major concern in the world. It is commonly associated with such things as smoking, obesity, poor diet, physical lack of activity, and lifestyle habits such as smoking, high cholesterol (hyperlipidemia), and high blood sugar (hyperglycemia). This disease is a significant health risk to the general population, and its treatment and prevention should remain one of the priorities.

II. LITERATURE SURVEY

"Predicting Diabetes Mellitus with Machine Learning Techniques:"

This paper examines the difficulty of making the right diagnosis of patients with diabetes mellitus. By apply ML models (SVM, RF, NB, eXtreme Gradient Boosting and DNN) to the PIMA Indian Diabetes and NHANES 19992016 dataset. These models involve real-life and on-line medical data that available. In experiment on the PIMA Indian Diabetes dataset, we achieve 79 percent accuracy of the RF model at binary classification when using a 60:40 train-test split in which features were selected using BORUTA. Conversely, XGBoost model performs well on the NHANES 19992016 data with an accuracy of 92% on binary classification & 91% on multiclass classification.

"The Role of Artificial Intelligence in Diabetes Research, Diagnosis and Prognosis: A Narrative Review:"

Artificial intelligence has proven useful, the ability of AI to handle a large amount of data allows it to arrive at significant conclusions. ML and DL are the most widespread technologies and are so much advanced due to the quicker computers and additional computing power. The paper provides a summary of the studies conducted concerning the use of AI to manage diabetes. Research in the well-known journals proved that AI can be used to possibly locate various species of diabetes with the assistance of bigdata and complicated formulas to assist physicians in making superior decisions. It also discusses the medical implications of AI in the diagnosis and ranking of diabetics as it will help in their earlier detection, fewer errors during their diagnosis and provide better results to patients. As is the case with any other sector, AI is limited. It is necessary to know these boundaries to make good use of AI in the treatment of diabetes.

"Predicting diabetes with multivariate analysis an innovative KNN-based classifier approach":

Diabetes appears to be severe chronic illness. People who are diabetic experience a long-term high level of blood glucose in their system due to the lack of responsiveness of cells to hormones. Such situations also result in prolonged issues or more health concerns. K-Nearest Neighborhood is an efficient user-friendly ML tool to create models that determine the probability

DOI: 10.62647/IJITCE2025V13I4PP370-374

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 4, 2025

of a person becoming ill, depending on the appropriate clinical data. To achieve objective, we designed a neuro-fuzzy inference K-Nearest Neighborhood (AF-KNN) learning-dependent predicting system, which predicts the behavior of patients in several various ways. This is a process of identifying the most appropriate set of neighborhoods where the probability of being mistaken is very low such that the final system is more able to predict the future.

"Association of type2 diabetes with coronary risk factors, clinical presentation, angiography, coronary interventions and follow-up outcomes: A single centre prospective registry:"

Background and goals A prospective study was conducted to determine how coronary artery disease (CAD) hospitalizations of patients with and without diabetes, the kind of care they receive, and the results that follow. Patients who underwent PCI who had primarily acute coronary syndromes were recruited one at a time between January 2018 and March 2021. These subjects were compared to diabetes and no diabetes patients to determine whether any changes occurred in their clinical and angiographic characteristics and results. The primary outcome was heart failure death and the large adverse cardiovascular events such as heart failure death, myocardial infarction, revascularization and stroke were significant. We identified the 95 percent confidence intervals (CI) and the Cox-proportional hazard ratios (HR). Findings there were 5181 individuals registered, 4139 men, and 1042 women, 4 917 individuals were experiencing acute cardiac syndrome (ACS), and 87% individuals with diabetes were diabetic. Diabetes was linked to more deaths from heart disease (HR 2.38, CI 1.135.02) and all causes (HR 1.85, CI 1.063.22).

"Identifying Reasons for Statin Nonuse in Patients with Diabetes Using Deep Learning of Electronic Health Records":

Background Statins are medications, which are recommended by guidelines aiming at reducing the risk of heart complications in individuals with diabetes. Identifying the factors that lead to individuals not taking their statins, which are likely to be available in unstructured electronic health record data. We intended to apply a DL technique to discover the reasons why not all diabetic individuals take their statins. Methods and Outcomes Between 2014 and 2020, adult patients with diabetes who lacked statin orders were identified using an electronic health record cohort of Northern California and included patients of various races and places. Out of unstructured electronic health record data, we applied a common DL, NLP system (Clinical Bidirectional Encoder Representations from Transformers) to identify individuals who did not take statins and reasons why they did not take statins. Of thirty-three thousand individuals with diabetes (15,580) the number of individuals who had no statin prescriptions was 14,000. Most of these individuals were women (49%), whites (36%), Asians (24%), and Hispanics (15%).

III. PROBLEM STATEMENT

- Diabetes is a chronic illness that may cause severe issues as cardiovascular diseases, renal failure and neuropathy. The impact of these complications on the quality of life of people can be tremendous.
- Some of the factors that have led to the increase in diabetes include lack of physical activities, poor diet, genes and the increased cases of obesity.
- One out of every three individuals with diabetes possesses health issues, economic issues and emotional issues.
- It increases health care expenses, reduces the productivity of the workers, strains health services, which all impact the economy and health of the society at large.

To address this critical issue, we propose a novel Ensemble DL clinical decision support system in which numerous datasets and algorithms will be used to make diabetes predictions more precise and enable those who are at a risk to obtain assistance when they require it.

IV. PROPOSED SYSTEM

The proposed system is a novel Ensemble Deep Learning clinical decision support tool designed to be able to predict diabetes correctly. The three health care data that are integrated are the "Pima Indian Diabetes Dataset (PIMA-IDD-I), the Diabetes Dataset of the Frankfurt Hospital in Germany (DDFH-G), and the Iraqi Diabetes Patient Dataset (IDPD-I)".

It employs a feature extraction algorithm with the ETC to obtain useful models in the datasets. Some of the methods, which are employed in the EDL framework to enhance the performance of the predictions based on various learning strategies, include ANN, CNN, LSTM, Stack-LSTM, Stack-ANN, Stack-CNN, and CNN + LSTM. Combining these algorithms, the system will be hoping to identify complex patterns and relationships in the data. This will ensure that predictions of diabetes are more accurate. The end result is that this approach will make it easier to have those who are at risk of developing diabetes and diagnosing them properly at an earlier period, and finding an effective approach through which they can be assisted.

Key Features of the Proposed System

- The proposed system integrates multiple datasets, enhancing model robustness and generalizability across diverse populations, allowing for a more

DOI: 10.62647/IJITCE2025V13I4PP370-374

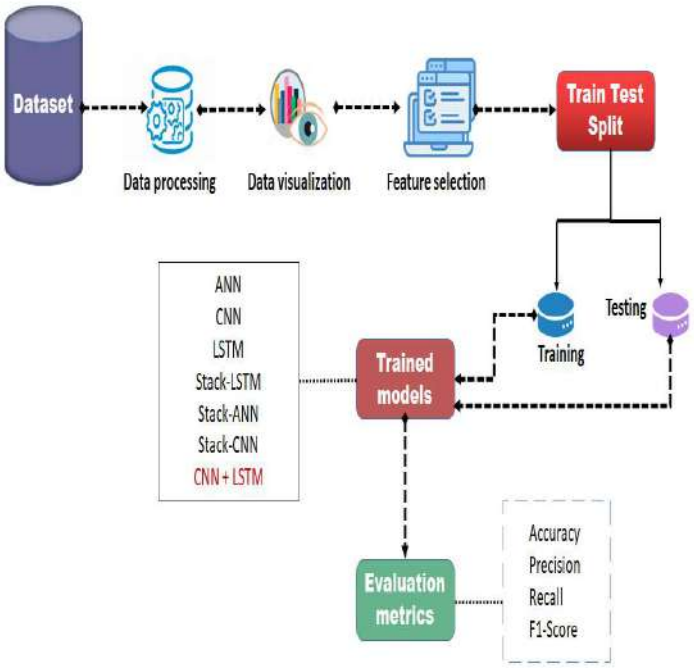
International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 4, 2025

comprehensive understanding of diabetes predictors and their interactions in varied clinical contexts.

- Employing an Ensemble Deep Learning framework enables the model to leverage the strengths of different algorithms, improving predictive performance by capturing complex data patterns and enhancing accuracy in diabetes prediction over traditional methods.
- The feature extraction using Extra Tree Classifier identifies the most relevant predictors, improving model efficiency by reducing dimensionality and focusing on critical features, which leads to better interpretability and clinical applicability.
- Combining various DL techniques, such as ANN, CNN, and LSTM, allows the system to utilize diverse learning strategies, improving predictive performance and facilitating early diagnosis and intervention for individuals at risk of developing diabetes



System Architecture

- Data loading: using this module we are going to import the dataset.
- Data Preprocessing: It involves cleaning, transforming, and organizing diabetes datasets to ensure quality and consistency. This step prepares the data for analysis, enhancing the accuracy of predictions.
- Data Visualization: presents processed data through graphical representations, such as charts and graphs.
- Extra Tree Feature Selection: It employs an ensemble method to identify and rank important features within diabetes datasets. By reducing dimensionality, this technique enhances model performance and simplifies the predictive analysis process.
- Splitting data into train & test: using this module data will be divided into train & test
- Model generation: Model building - ANN - CNN - LSTM - Stack-LSTM - Stack-ANN - Stack-CNN - CNN + LSTM {Pima-IDDI, DDFH-G & IDPD-I}. Performance evaluation metrics for each algorithm is calculated.
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

V. CONCLUSION

Lastly, the proposed Ensemble DL clinical decision support system to predict diabetes has much potential as far as making predictions more precise and enhancing the management of patients in healthcare. Having viewed numerous other algorithms, the CNN + LSTM hybrid model appeared to be a high-performing algorithm, which incorporates the most effective attributes of both CNN and LSTM networks. With this integration, the model is able to consider the space-based features as well as the time-based relationships

DOI: 10.62647/IJITCE2025V13I4PP370-374

in the diabetes datasets. Through CNN to extract features and LSTM to learn sequences, the hybrid model addresses the complexities of diabetes prediction, enabling accurate and timely assessments. Ultimately, the implementation of high-performance algorithms like CNN + LSTM within the clinical decision support system holds the promise of better diabetes management, improved patient outcomes, and a proactive approach to healthcare intervention. This innovative system can significantly contribute to reducing the impact of diabetes on global health.

REFERENCES

[1]. H. L. Tong, H. Ng, and H. Arul Ananthan, ‘‘Predicting diabetes mellitus with machine learning techniques,’’ J. Eng. Technol. Appl. Phys., vol. 6, no. 1, pp. 91–99, Mar. 2024.

[2]. S. Binhowemel, M. Alfakhri, K. AlReshaid, and A. Alyani, ‘‘Role of artificial intelligence in diabetes research diagnosis and prognosis: A narrative review,’’ J. Health Inform. Developing Countries, vol. 17, no. 2, pp. 1–12, Aug. 2023. Available: https://www.jhidc.org/index.php/jhidc/article/view/410

[3]. B. V. V. S. Prasad, S. Gupta, N. Borah, R. Dineshkumar, H. K. Lautre, and B. Mouleswararao, ‘‘Predicting diabetes with multivariate analysis an innovative KNN-based classifier approach,’’ Preventive Med., vol. 174, Sep. 2023, Art. no. 107619, doi: 10.1016/j.ypmed.2023.107619.

[4]. R. Gupta, S. Lodha, K. K. Sharma, S. K. Sharma, J. S. Makkar, A. Bana, V. Natani, S. Kumar, S. Bharati, and S. K. Sharma, ‘‘Association of type 2 diabetes with coronary risk factors, clinical presentation, angiography, coronary interventions and follow-up outcomes: A single centre prospective registry,’’ Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 17, no. 2, Feb. 2023, Art. no. 102709, doi: 10.1016/j.dsx.2023.102709.

[5]. Mohammed Abdul Bari , Arshad Ahmad Khan Mohammad,” Securing MANETs Against Black Hole Attacks: A Novel Approach Using Secure Knowledge Algorithm-International Conference on Advances in Emerging Trends in Computer Application (ICAETC-2023)-Scopus 21-22 Dec 2023.

[6]. M.A.Bari & Shahanawaj Ahamad, “Managing Knowledge in Development of Agile Software”, in International Journal of Advanced Computer Science & Applications (IJACSA), ISSN: 2156-5570, Vol: 2, No: 4, pp: 72-76, New York, U.S.A., April 2011

[7]. A. Sarraju, A. Zammit, S. Ngo, C. Witting, T. Hernandez-Boussard, and F. Rodriguez, ‘‘Identifying reasons for statin nonuse in patients with diabetes using deep learning of electronic health records,’’ J. Amer. Heart Assoc., vol. 12, no. 7, Apr. 2023, Art. no. e028120.

[8]. R. M. Alamoudi et al., ‘‘Fasting Ramadan in patients with T1DM— Saudi Arabia versus other countries during the COVID-19 pandemic,’’ Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 17, no. 1, Jan. 2023, Art. no. 102676, doi: 10.1016/j.dsx.2022.102676.

[9]. World Population Rev. Diabetes Rates By Country 2023. Accessed: Feb. 10, 2023. [Online]. https://worldpopulationreview.com/country-rankings/diabetes-rates-by-country

[10]. V. Jaiswal, A. Negi, and T. Pal, ‘‘A review on current advances in machine learning based diabetes prediction,’’ Primary Care Diabetes, vol. 15, no. 3, pp. 435–443, Jun. 2021 doi: 10.1016/j.pcd.2021.02.005.

[11]. [9] A. Mujumdar and V. Vaidehi, ‘‘Diabetes prediction using machine learning algorithms,’’ Proc. Comput. Sci., vol. 165, pp. 292–299, Jan. 2019.

[12]. J. Chaki, S. T. Ganesh, S. K. Cidham, and S. A. Theertan, ‘‘Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review,’’ J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 6, pp. 3204–3225, Jun. 2022, doi: 10.1016/j.jksuci.2020.06.013.

[13]. R. K. Nadesh and K. Arivuselvan, ‘‘Type 2: Diabetes mellitus prediction using deep neural networks classifier,’’ Int. J. Cogn. Comput. Eng., vol. 1, pp. 55–61, Jun. 2020
doi: 10.1016/j.ijcce.2020.10.002.

[14]. H. Lai, H. Huang, K. Keshavjee, A. Guergachi, and X. Gao, ‘‘Predictive

DOI: 10.62647/IJITCE2025V13I4PP370-374

ISSN 2347–3657

Volume 13, Issue 4, 2025

models for diabetes mellitus using machine learning techniques,'' BMC Endocrine Disorders, vol. 19, no. 1, pp. 1–9, Dec. 2019. doi: 10.1186/s12902-019-0436-6.

[15]. M. Bhattacharya and D. Datta, ''Development of predictive models of diabetes using ensemble machine learning classifier,'' in Proc. 1st Int. Conf. Advancements Smart Comput. Inf. Secur. (ASCIS), Rajkot, India, 2022, pp. 377–388.

[16]. Y. Du, A. R. Rafferty, F. M. McAuliffe, L. Wei, and C. Mooney, ''An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus,'' Sci. Rep., vol. 12, no. 1, p. 1170, Jan. 2022, doi: 10.1038/s41598-022-05112-2.

[17]. O. A. Ebrahim and G. Derbew, ''Application of supervised machine learning algorithms for classification and prediction of type-2 diabetes disease status in afar regional state, Northeastern Ethiopia 2021,'' Sci. Rep., vol. 13, no. 1, p. 7779, May 2023, doi: 10.1038/s41598-023-34906-1.

DOI: 10.62647/IJITCE2025V13I4PP370-374