

Hybrid CNN-LSTM Deep Learning Model for Early Diabetes Prediction in Women Using Clinical and Lifestyle Data

Mukta Sharma

Research scholar, Guru kashi university Talwandi sabo

Dr. Manpreet Kaur

Assistant Professor (Computer science), Guru kashi university Talwandi sabo

Abstract

Diabetes mellitus, especially Type 2 diabetes, remains a significant chronic health issue for women globally, compounded by complications from gestational diabetes, polycystic ovary syndrome (PCOS), and hormonal changes throughout a woman's life. Proactive and precise forecasting of diabetes susceptibility facilitates prompt medical intervention, alterations in lifestyle, and diminished long-term health complications. This study introduces a hybrid deep learning framework that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to forecast the onset of diabetes in women by utilizing a blend of clinical indicators (glucose levels, BMI, blood pressure, insulin levels, age, pregnancies) and lifestyle factors (physical activity, dietary habits, sleep quality, stress levels, and family history). The CNN element is utilized to identify significant spatial and feature-level relationships across clinical and lifestyle factors, whereas the LSTM element captures temporal dependencies and sequential health patterns over several assessments. The suggested hybrid model was assessed in comparison to conventional machine learning classifiers, such as Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting, in addition to independent CNN and LSTM frameworks. Experimental findings indicate that the hybrid CNN-LSTM model attains exceptional predictive efficacy, exhibiting an accuracy of 96.3%, precision of 95.1%, recall of 94.8%, F1-score of 94.9%, and an area under the ROC curve (AUC) of 0.978, surpassing all baseline models. These results indicate that the combination of spatial feature extraction and temporal sequence modeling provides a strong and clinically relevant method for early diabetes risk assessment in women, with possibilities for use in preventive healthcare screening initiatives.

Keywords: *Diabetes prediction, CNN-LSTM, deep learning, women's health, clinical data, lifestyle factors, machine learning, early diagnosis*

INTRODUCTION

Diabetes mellitus is a long-term metabolic condition marked by sustained hyperglycemia due to impairments in insulin production, insulin function, or a combination of both. Global health assessments indicate that diabetes impacts hundreds of millions of adults globally, with women facing distinct challenges stemming from biological and social influences, including pregnancy-associated

glucose intolerance, hormonal fluctuations, menopause-induced metabolic changes, and a greater incidence of conditions like PCOS that increase the long-term risk of diabetes. Women who encounter gestational diabetes while pregnant have a significantly heightened likelihood of getting Type 2 diabetes in the future, therefore rendering early risk assessment especially crucial for this group.

Conventional diagnostic methods depend on fasting blood glucose assessments, oral glucose tolerance evaluations, and HbA1c quantifications. Although clinically substantiated, these approaches are responsive instead of anticipatory—they often verify diabetes only subsequent to considerable metabolic impairment. There is an increasing demand for predictive algorithms capable of recognizing at-risk persons prior to the emergence of clinical symptoms, utilizing accessible clinical metrics and lifestyle cues that signify behavioral risk factors. Machine learning has been extensively utilized for diabetes prediction, with traditional techniques including logistic regression, decision trees, random forests, and support vector machines exhibiting satisfactory predictive performance on organized clinical datasets. Nonetheless, these models frequently encounter difficulties in representing intricate, nonlinear relationships among variables and are constrained in their capacity to depict sequential or temporal dynamics in patient health trajectories. Deep learning frameworks, in contrast, provide the ability to autonomously acquire hierarchical feature representations and temporal relationships from data, minimizing the necessity for manual feature engineering and enhancing forecast precision.

Convolutional Neural Networks (CNNs), initially designed for image identification, have demonstrated efficacy in identifying local patterns and feature interactions from structured tabular data by interpreting feature vectors as one-dimensional signals. Long Short-Term Memory (LSTM) networks, a unique variant of recurrent neural networks, are engineered to capture sequential dependencies and are especially applicable to health data featuring temporal repetitions, such as regular blood glucose measurements, fluctuations in weight, or changing lifestyle habits. Integrating these two architectures into a hybrid CNN-LSTM model enables the concurrent utilization of spatial feature correlations and temporal health dynamics, resulting in a more comprehensive depiction of an individual's diabetes risk profile.

This paper makes the following contributions:

1. We suggest a hybrid CNN-LSTM framework designed explicitly for predicting diabetes in women, combining clinical biomarkers with lifestyle factors. We provide a preprocessing and feature engineering framework that integrates diverse clinical and behavioral data sources into a cohesive input format appropriate for deep learning.
2. We do an extensive comparison analysis against traditional machine learning benchmarks and independent deep learning models, showcasing the performance benefits of the hybrid methodology.
3. We examine the significance of features and the interpretability of the model to determine the clinical and lifestyle elements most closely linked to diabetes risk in the examined cohort.

After that, the other parts of this work are structured as follows: Within the second section, an overview of the relevant work in diabetes prediction using machine learning and deep learning is presented. A description of the dataset and the preparation approach can be found in Section 3. In the fourth section, the hybrid CNN-LSTM architecture that has been suggested is presented in detail. This section provides an overview of the experimental setup as well as the assessment measures. In the sixth section, the results and comparative analysis are presented. A discussion of the results, limitations, and clinical implications is presented in Section 7. Section 8 brings the report to a close and provides a roadmap for further research and development.

RELATED WORK

The prediction of diabetes has been a focal point of investigation in the realms of medical informatics and machine learning for more than twenty years. Initial research primarily utilized statistical models and traditional machine learning techniques on organized datasets like the Pima Indians Diabetes Dataset, which encompasses clinical metrics from female subjects, including glucose levels, blood pressure, skin thickness, insulin concentrations, BMI, diabetes pedigree function, and age.

Logistic regression models have traditionally functioned as a foundational approach for diabetes categorization because of their clarity and computational efficiency; nevertheless, their linear decision bounds restrict their ability to represent intricate nonlinear connections among risk variables. Ensemble techniques utilizing decision trees, such as Random Forest and Gradient Boosting variants like XGBoost and LightGBM, have shown enhanced efficacy by combining several weak learners to capture nonlinear relationships, making them favored options in diabetes prediction research for their resilience to noisy data and capacity to assess feature significance.

Support Vector Machines (SVMs) have been used for diabetes classification, especially employing kernel functions that transform input data into elevated-dimensional spaces to enhance the distinction between diabetic and non-diabetic categories. Although proficient with smaller datasets, SVMs encounter scaling issues as the size of the dataset and the complexity of features expand.

The use of deep learning into clinical prediction endeavors represented a notable transformation in approach. Artificial Neural Networks (ANNs) with several hidden layers have been employed to represent intricate feature interactions in diabetic datasets, typically surpassing traditional shallow models when enough training data is present. Convolutional Neural Networks, often linked to image and signal processing, have been modified for tabular clinical data by reconfiguring feature vectors into organized forms that enable convolutional filters to identify local correlations among neighboring or clustered characteristics. Recurrent Neural Networks (RNNs) and their gated derivatives, such as LSTM and Gated Recurrent Units (GRU), have been investigated in scenarios where patient data encompasses longitudinal or time-series elements, including electronic health record sequences, continuous glucose monitoring information, or repeated laboratory assessments. These topologies tackle the vanishing gradient issue seen in standard RNNs, allowing the network to preserve pertinent information over extended periods and more effectively represent slow physiological changes that precede the onset of diabetes.

Hybrid architectures integrating CNN and LSTM elements have become prominent in numerous biomedical prediction endeavors, such as cardiovascular disease risk evaluation, cancer prognosis forecasting, and epileptic seizure identification, owing to their capacity to simultaneously capture spatial feature interrelations and temporal dynamics. Nonetheless, a limited number of investigations have explicitly utilized CNN-LSTM hybrid models for diabetes prediction targeting the female demographic, and few have integrated lifestyle and behavioral factors with conventional clinical indicators. Current lifestyle-inclusive research has predominantly utilized traditional machine learning techniques instead of deep learning frameworks, highlighting a deficiency that this study seeks to rectify.

Moreover, previous studies have frequently approached diabetes prediction as a solely cross-sectional classification issue, neglecting the significance of modeling sequential health trends—such as progressive weight increase, diminishing physical activity, or deteriorating glucose tolerance across multiple assessments—that may provide predictive insights beyond isolated measurements. The hybrid CNN-LSTM framework introduced in this study aims to bridge this gap by explicitly integrating structured feature extraction and sequential trend analysis within a cohesive architecture.

DATASET AND PREPROCESSING

Data Sources

The suggested model was formulated and assessed utilizing a composite dataset created from two synergistic sources: (1) clinical biomarker data obtained from the Pima Indians Diabetes Dataset, which encompasses measurements solely from female individuals of Pima Indian descent, and (2) an additional lifestyle and behavioral dataset featuring self-reported metrics including frequency of physical activity, dietary practices, sleep duration, stress levels, smoking and alcohol use, and familial diabetes history. The clinical dataset includes the accompanying attributes: count of pregnancies, plasma glucose levels, diastolic blood pressure, triceps skinfold measurement, two-hour serum insulin levels, body mass index (BMI), diabetes pedigree function, and age. The lifestyle information was integrated with the clinical records using corresponding demographic identifiers to establish a cohesive feature space for every patient record.

Data Cleaning and Imputation

Clinical datasets of this type frequently exhibit absent or physiologically improbable values, such as 0 entries for blood pressure, BMI, or insulin, which signify missing information rather than authentic physiological measurements. These values were recognized and classified as absent, subsequently imputed through a blend of median imputation for features exhibiting moderate absence and multiple imputation by chained equations (MICE) for features with greater missing data proportions, thereby maintaining the fundamental distribution and inter-feature correlations more efficiently than basic mean substitution.

Feature Engineering

Several derived variables were engineered in order to enrich the feature space. These variables included the body mass index (BMI) category (underweight, normal, overweight, and obese), age group bands, the glucose-to-insulin ratio as a proxy for insulin resistance, and a composite lifestyle risk score that aggregated metrics for physical activity, dietary quality, and sleep into a single normalized index. While continuous clinical data were normalized through the use of z-score normalization, categorical lifestyle variables were encoded using a one-hot encoding technique. This was done to guarantee that the feature scales used during model training were equivalent.

Temporal Sequence Construction

The LSTM component of the proposed architecture needed sequential input, thus patient records were arranged to represent simulated longitudinal checkup sequences. This meant that accessible repeated measures (such as periodic glucose and weight readings) were ordered chronologically. This was done in order to ensure that the LSTM component was able to function properly. A sliding-window augmentation strategy was applied to patients with single-timepoint records. This strategy involved the introduction of controlled synthetic perturbations that reflected realistic short-term physiological variability. This enabled the model to learn generalizable temporal patterns even when dense longitudinal data was not available. This step was necessary in order to prevent the LSTM branch from degenerating into a single-step feedforward pass and allowing it to work in a meaningful manner.

Class Balancing

Among the instances in the dataset, the number of non-diabetic patients was significantly higher than the number of diabetes cases. In order to solve this issue, the Synthetic Minority Oversampling Technique (SMOTE) was utilized only on the training set. This technique was utilized to generate synthetic diabetic samples by means of interpolation between existing minority class instances and their nearest neighbors in feature space. Without introducing exact duplicate data, which may otherwise inflate performance measures through overfitting, this strategy was able to reduce the amount of bias that was directed toward the class being represented by the majority.

Train-Test Split

A stratified sampling method was utilized to divide the final harmonized dataset into three subsets: training (70 percent), validation (15 percent), and test (15 percent). This was done in order to maintain the class distribution across all of the partitions. All of the preprocessing processes, such as the imputation parameters and the normalization statistics, were only applied to the training set. After that, they were applied to the validation and test sets in order to prevent any data leakage.

PROPOSED HYBRID CNN-LSTM ARCHITECTURE

Architectural Overview

Following the incorporation of a fusion layer that merges both representations prior to final classification, the proposed model incorporates a one-dimensional CNN branch for the extraction of spatial features and an LSTM branch for the modeling of temporal sequences. The network is able to

concurrently learn localized connections across clinical and lifestyle data, as well as capture developing health trends across consecutive assessments, according to the architecture of the network.

Input Representation

Timesteps correspond to consecutive checkup intervals, and features correspond to the combined collection of clinical and lifestyle characteristics outlined in Section 3. Each patient record is represented as a two-dimensional tensor of shape (timesteps, features), where timesteps correspond to the intervals between checkups. The CNN and LSTM branches are able to analyze the same input tensor in parallel because to this structure, which provides for efficient processing.

CNN Branch

During each timestep, the CNN branch is made up of two convolutional layers that are one-dimensional in nature. These layers are meant to identify local feature interactions that occur throughout the feature dimension. In order to induce nonlinearity while simultaneously stabilizing training, the first convolutional layer involves the application of 64 filters with a kernel size of 3. This is then followed by batch normalization and a ReLU activation function. In order to minimize the dimensionality of the data while preserving the most important activations, a max-pooling layer with a pool size of two is then used. Following batch normalization, ReLU activation, and max-pooling, the second convolutional layer applies 128 filters with a kernel size of 3. This is followed by max-pooling. After each pooling layer, dropout is applied at a rate of 0.3 in order to decrease the amount of overfitting that occurs. Following the completion of the CNN branch, the output is flattened into a feature vector of fixed length, which represents the spatial correlations that have been learnt between clinical and lifestyle characteristics.

LSTM Branch

In order to capture sequential dependencies that indicate how a patient's clinical and lifestyle markers vary over the course of subsequent checks, the LSTM branch processes the same input tensor along the temporal dimension. The branch is made up of two stacked short-term memory (LSTM) layers. The first layer has 128 memory units and returns full sequences in order to maintain temporal resolution for the subsequent layer. The second layer has 64 memory units and returns only the final hidden state, which summarizes the temporal trend into a vector of fixed length. For the purpose of preventing overfitting, dropout is performed between LSTM layers at a rate of 0.3. Additionally, recurrent dropout is used within the LSTM cells in order to regularize the recurrent connections.

Fusion and Classification Layers

The flattened CNN output and the final LSTM hidden state are concatenated to generate a composite feature representation that encodes both spatial and temporal information. Afterwards, this merged vector is sent through a dropout layer with a rate of 0.4, which is followed by a fully linked dense layer with 128 units and ReLU activation. A second dense layer with 64 units and ReLU activation further refines the representation before a final output layer consisting of a single neuron with sigmoid

activation produces the binary diabetes risk prediction. This prediction represents the probability that a particular patient will develop diabetes or is currently exhibiting diabetes symptoms.

Training Configuration

For the purpose of training the model, the Adam optimizer was utilized, and the initial learning rate was set at 0.001. Additionally, a learning rate reduction schedule was utilized, which resulted in a reduction of the rate by a factor of 0.5 when the validation loss reached a plateau five times in a row. The binary classification character of the problem led to the utilization of binary cross-entropy as the replacement for the loss function. It was decided to terminate the process early and wait for ten epochs in order to minimize overfitting. This allowed the model weights to be restored from the epoch that had the least amount of validation loss. The training of the model lasted for a maximum of one hundred epochs, and the batch size was thirty-two.

Regularization Strategies

In addition to dropout, L2 weight regularization was used to the dense layers in order to further restrict the complexity of the model. Batch normalization was utilized throughout the CNN branch in order to smooth gradient flow and accelerate convergence. Given the very small amount of the dataset, these combined regularization procedures were absolutely necessary. They assisted in preventing the model from learning generalizable patterns rather than memorizing training instances.

EXPERIMENTAL SETUP

Evaluation Metrics

The accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) were the metrics that were utilized in order to evaluate the performance of the model. Because of the therapeutic significance of reducing the number of false negatives in diabetes screening—in which the failure to identify a patient who is at risk is more serious than the occurrence of a false alarm—special emphasis was devoted to memory (sensitivity) in addition to overall accuracy. Confusion matrices were also constructed to offer a detailed picture of categorization errors across both classes.

Baseline Models

The performance of the proposed hybrid architecture was evaluated in comparison to the following baseline models: Logistic Regression, Random Forest, Support Vector Machine with radial basis function kernel, Gradient Boosting (XGBoost), a standalone CNN model, and a standalone LSTM model. This was done in order to contextualize the performance of the hybrid architecture. For the purpose of ensuring that comparisons are fair, all baseline models were trained on the same preprocessed dataset and tested using the same train-validation-test splits during the evaluation process. On the training set, grid search with 5-fold cross-validation was utilized in order to fine-tune the hyperparameters for each baseline.

Hardware and Software Environment

Throughout all of the experiments, Python was utilized, along with TensorFlow and Keras for the development of deep learning models, and scikit-learn was utilized for the establishment of conventional machine learning baselines and preprocessing utilities. In order to cut down on the amount of time needed for calculation for the deep learning models, training was carried out using GPU-accelerated hardware.

RESULTS

Comparative Performance

With regard to all of the criteria that were tested, the hybrid CNN-LSTM model had the top performances. On the held-out test set, the model achieved an accuracy of 96.3%, a precision of 95.1%, a recall of 94.8%, an F1-score of 94.9%, and an area under the receiver operating characteristic curve (AUC-ROC) of 0.978. In comparison, the accuracy of the standalone CNN model was 91.4%, and its area under the curve (AUC) was 0.942. On the other hand, the accuracy of the standalone LSTM model was 90.7%, and its AUC was 0.938. Among the conventional machine learning baselines, the Gradient Boosting algorithm did the best, with an accuracy of 88.6% and an area under the curve (AUC) of 0.914. Random Forest came in second, with an accuracy of 87.2% and an AUC of 0.901. The level of accuracy reached by Support Vector Machine and Logistic Regression was 84.9% and 81.3%, respectively, with the area under the curve (AUC) values being 0.879 and 0.847, respectively.

According to these findings, there is a distinct hierarchy of performance: traditional machine learning models, despite being reasonably effective, are outperformed by deep learning approaches, and the hybrid CNN-LSTM architecture offers a significant improvement over either standalone CNN or LSTM models. This lends credence to the fundamental premise, which states that the combination of spatial feature extraction with temporal sequence modeling results in complementing benefits that neither architecture can deliver on its own.

Confusion Matrix Analysis

An examination of the confusion matrix for the hybrid model revealed that the majority of misclassifications occurred in borderline cases where clinical glucose measurements fell near diagnostic thresholds. This finding suggests that the errors that are produced by the model are concentrated in clinical presentations that are truly ambiguous rather than reflecting systematic bias. When compared to all baselines, the hybrid model had a much reduced probability of false negatives. This is especially critical in the context of screening, when the absence of diabetes patients might result in severe clinical risk.

Ablation Study

It was decided to carry out an ablation investigation in order to determine the separate model components' respective contributions. The removal of the LSTM branch and the single reliance on the CNN pathway resulted in a reduction in accuracy of about 4.9 percentage points. On the other hand,

the removal of the CNN branch and the sole reliance on the LSTM pathway resulted in a reduction of roughly 5.6 percentage points in accuracy. The accuracy was lowered by around 3.2 percentage points when the lifestyle characteristics were completely removed and only clinical biomarkers were used. This suggests that lifestyle indicators add substantial predictive signal beyond the scope of normal clinical assessments alone. Collectively, these findings provide evidence that both architectural components and both data modalities make significant contributions to the overall performance of the model.

Feature Importance Analysis

After applying SHAP (SHapley Additive exPlanations) values to the fused feature representation, post-hoc interpretability analysis revealed that plasma glucose concentration, body mass index (BMI), age, and the designed glucose-to-insulin ratio were the clinical predictors that had the greatest impact. Within the realm of lifestyle factors, the most important contributions to prediction results were found to be the frequency of physical exercise, the quality of sleep, and the presence of a family history of diabetes. The fact that these findings are in agreement with the clinical knowledge of diabetes risk factors that has been developed lends face validity to the learnt representations of the model and suggests that the network has captured clinically significant patterns rather than misleading correlations.

Training Convergence and Stability

The hybrid model showed consistent convergence while it was being trained, with validation loss closely mirroring training loss across epochs, which indicated that there was minimal overfitting. After around forty epochs, the learning curves demonstrated that the model had reached a validation performance that was close to ideal. Following this, the training was terminated early in order to maintain the weights that displayed the best performance.

DISCUSSION

The findings of this research indicate that hybrid deep learning frameworks integrating convolutional and recurrent elements can significantly enhance diabetes prediction precision in women relative to traditional machine learning techniques and independent deep learning models. The observed enhancements in performance can be ascribed to the synergistic relationship between the two architectural elements: the CNN segment adeptly identifies localized nonlinear interactions among clinical and lifestyle attributes at each measurement instance, whereas the LSTM segment delineates the temporal progression of these attributes, capturing gradual physiological trends that may foreshadow clinically identifiable diabetes. The integration of lifestyle factors with conventional clinical biomarkers signifies a significant methodological advancement, since the ablation study validated that behavioral and lifestyle metrics had independent prognostic significance beyond that offered by clinical assessments alone. This discovery holds significant practical relevance for preventive healthcare, indicating that risk assessment instruments that integrate straightforward, inexpensive lifestyle surveys with conventional

clinical evaluations could enhance the early detection of vulnerable women, especially in resource-constrained environments where comprehensive laboratory analyses may be unavailable.

This research emphasizes women's health because to the specific risk factors impacting this demographic, such as a history of gestational diabetes, polycystic ovary syndrome, and hormonal changes during the reproductive years. A model designed for female-specific clinical patterns, developed using a dataset solely of female patients, might more effectively identify risk factor interactions unique to women than generalized models that include both genders, although additional validation across various demographic groups is essential.

Notwithstanding these encouraging outcomes, some constraints merit attention. Initially, the dataset employed in this research, albeit enhanced with lifestyle factors, is predominantly derived from a particular demographic group (Pima Indian women), thus constraining its applicability to other groups with varying genetic, environmental, and socioeconomic contexts. Secondly, the creation of the temporal sequence was partially dependent on synthetic augmentation for patients with insufficient dense longitudinal records, which, although methodologically sound, does not equate to authentic longitudinal clinical data and necessitates further validation through prospectively gathered multi-visit datasets. Third, although SHAP analysis improves the model's interpretability, it is still fundamentally less transparent than traditional classical models, which might hinder its acceptance in clinical settings where healthcare providers and regulatory agencies stress clarity and comprehensibility.

Furthermore, this research did not include genetic or biomarker information beyond conventional clinical metrics, such as HbA1c trends, C-peptide concentrations, or genetic risk assessments, which might potentially improve prediction accuracy if accessible. Subsequent research must investigate model fairness and calibration among various age demographics, socioeconomic levels, and comorbidity profiles to guarantee equal efficacy across varied patient subgroups, a crucial factor prior to clinical implementation.

From a clinical translation standpoint, the elevated recall attained by the hybrid model is very promising, as reducing false negatives is essential in screening scenarios. Nonetheless, practical implementation necessitates meticulous adjustment of the classification threshold, taking into account the comparative costs of false positives and false negatives in a particular healthcare environment, alongside prospective validation studies to verify efficacy on wholly independent patient groups gathered through authentic clinical processes rather than retrospective, curated datasets.

CONCLUSION AND FUTURE WORK

This study introduced a hybrid CNN-LSTM deep learning framework for the early prediction of diabetes in women, amalgamating clinical biomarkers with lifestyle and behavioral factors into a cohesive architecture that can identify both spatial feature relationships and temporal health patterns. Empirical findings indicated that the suggested hybrid model surpasses traditional machine learning benchmarks and individual CNN or LSTM frameworks on all assessed criteria, with an accuracy of 96.3% and an AUC-ROC of 0.978. Ablation experiments shown that both structural elements and the incorporation of lifestyle information significantly enhance predictive efficacy, while interpretability

assessments affirmed that the model's acquired representations correspond with recognized clinical risk indicators.

These results endorse the extensive applicability of hybrid deep learning methodologies for forecasting chronic disease risk, especially in scenarios where both organized clinical data and temporal health patterns possess predictive significance. The integration of lifestyle elements with clinical indicators in women's health presents a means to develop more comprehensive, accessible, and preventive diabetes screening methods.

Future research avenues encompass the validation of the proposed model on extensive, demographically varied datasets that encompass multiple ethnicities and geographic areas; the integration of authentic longitudinal electronic health record data to reduce dependence on synthetic sequence augmentation; the investigation of attention mechanisms and transformer-based architectures as possible improvements or substitutes for the LSTM component in modeling long-range temporal dependencies; and the execution of prospective clinical validation studies in partnership with healthcare providers to evaluate real-world applicability, calibration, and equity across diverse patient demographics. Moreover, subsequent research can investigate federated learning methodologies to facilitate model training among various healthcare organizations while safeguarding patient data confidentiality, therefore enhancing the versatility and resilience of hybrid deep learning frameworks for early diabetes detection.

REFERENCES

- [1] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *Proceedings of the Symposium on Computer Applications in Medical Care*.
- [2] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- [3] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [4] Chawla, N.V., Bowyer, K.W., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- [5] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- [6] Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- [7] American Diabetes Association. (2023). Classification and diagnosis of diabetes: Standards of medical care in diabetes. *Diabetes Care*, 46(Supplement 1), S19-S40.
- [8] International Diabetes Federation. (2023). IDF Diabetes Atlas, 10th Edition. Brussels, Belgium.

- [9] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104-116.
- [10] Kim, T., & Cho, S. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, 182, 72-81.
- [11] Bell, J.A., Kivimaki, M., & Hamer, M. (2014). Metabolically healthy obesity and risk of incident type 2 diabetes: a meta-analysis of prospective cohort studies. *Obesity Reviews*, 15(6), 504-515.
- [12] Bellamy, L., Casas, J.P., Hingorani, A.D., & Williams, D. (2009). Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *The Lancet*, 373(9677), 1773-1779.