Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 2, 2026

EDGE OF THINGS BASED DIABETES PREDICTION USING MACHINE LEARNING

Muhammad Touqeer Zahoor[1], Muhammad Safdar Amin Khan[*2]

[1]*Department of Computer Sciences Comsats University Islamabad*
[*2]*Department of Computer Sciences Alhamd Islamic University Islamabad*

DOI: https://doi.org/10.5281/zenodo.18517167

Keywords
Diabetes prediction; Edge of Things; Machine learning; Deep learning; Ensemble learning; Healthcare analytics; Class imbalance

Article History
Received: 08 December 2025
Accepted: 23 January 2026
Published: 07 February 2026

Corresponding Author: *
Muhammad Safdar Amin Khan

Abstract
Diabetes mellitus is a fast-increasing health issue in the world which needs to be diagnosed and managed in time to minimize complications and health expenses. Recent developments in machine learning have shown good prospects towards enhancing diabetes prediction, but, the majority of those existing solutions are based on centralized cloud models which are characterized by a great latency, privacy reasons, and inability to use them in resource-limited settings. This paper attempts to deal with these issues by suggesting an Edge of Things (EoT)-based diabetes predictive model based on the integration of hybrid deep learning models with ensemble machine learning algorithms to achieve decentralized and real-time prediction of diseases. The suggested structure will include extensive data preparation, features normalization, and the class imbalance to optimize predictive accuracy. Hybrid deep learning models are used to learn the complex nonlinear association among demographic, clinical and behavioral variables and ensemble learning methods exploit the synergistic advantages of multiple classifiers to enhance their robustness and generalization. The system will greatly decrease the response latency and bandwidth consumption, and will avoid relying on constant internet connection, which will also increase data privacy and security by deploying trained models on the edge. The experimental assessment of the publicly available diabetes data base shows that the hybrid-ensemble framework proposed is more accurate, sensitive, and stable as compared to the conventional one-model frameworks. The findings indicate that imbalance-conscious learning and edge-based intelligence is effective in healthcare analytics. On the whole, this research provides a solution to detecting diabetes at an early stage saving privacy and being scalable and efficient to support the proactive and individualized healthcare delivery in resource-restricted conditions.

INTRODUCTION

Diabetes mellitus can be considered as one of the most widespread and the fastest growing chronic illnesses on the planet and is a significant burden to the contemporary healthcare services. It is a metabolic condition, which includes sustained hyperglycemia caused by impairments in insulin release, insulin activity, or both (World Health Organization [WHO], 2023). The global health reports show that the population of people with diabetes has grown significantly over the last several decades, and the number of diabetic patients is expected to grow even more in the future because of population growth, age, and lifestyle changes (International Diabetes Federation [IDF], 2021). Long-term complications of diabetes are numerous in nature, and they may include cardiovascular diseases, renal failure, neuropathy, retinopathy, amputation of limbs, and early death (Zheng et al., 2018). The complications not only worsen the quality of life of patients but also cost the healthcare systems, families, and even the entire society a significant sum of money (Bommer et al., 2017). Consequently, timely diagnosis and proper management of diabetes is very essential in lowering morbidity, mortality and health costs.

Lifestyle and socio-economic factors have been closely associated with the rapid rise in the rates of diabetes. Urbanization and industrialization have resulted in less physical activities, more consumption of processed foods and more stress and this contributes to obesity and insulin resistance (Hu, 2011). Genetic predisposition is another factor, which is important especially in the case of interaction with environmental and behavioral factors (Alberti & Zimmet, 2013). The diabetes epidemic is unproportionately high in developing and underdeveloped countries because of the poor healthcare infrastructure, the absence of awareness, and access to preventive and diagnostic care (IDF, 2021). In most instances, they are not diagnosed until complications arise which are severe making both less effective and more expensive to treat (WHO, 2023). These issues demonstrate the emergent need of smart and scalable solutions that allow detecting and continuously monitoring diabetes during its early stages especially in resource-limited settings.

Innovations in information and communication technologies (ICT) in the recent past have brought a lot of changes in healthcare delivery and in the management of diseases. The utilization of electronic health records, wearable technologies, mobile health tools, and Internet of Things (IoT) technologies has facilitated data collection of patients in real-life environments (Islam et al., 2015). These technologies produce big data in heterogeneous form in the form of physiological signals, clinical measurements and information on lifestyles. Such data has enabled a paradigm shift in the conventional reactive healthcare paradigms to predictive, preventive, and personalized care (Raghupathi & Raghupathi, 2014). In this regard, the use of data-driven methods has gained more significance in clinical decision-making and health outcomes.

Machine learning (ML) has become an effective instrument of examining intricate and manifold healthcare information. In contrast to traditional statistical methods, the ML algorithms may automatically discover the patterns and relationships in data without them being explicitly programmed (Jordan and Mitchell, 2015). ML models have had remarkable results in an extensive variety of healthcare applications, such as disease diagnosis, risk prediction, treatment recommendation, and patient monitoring (Esteva et al., 2019). ML algorithms used in the prediction of diabetes consider clinical, demographic, and behavioral factors, including age, body mass index, blood glucose level, blood pressure, insulin levels, physical activity, and family history to predict the likelihood of a particular individual developing the disease (Kavakiotis et al., 2017). Research has found that the accuracy and sensitivity of prediction based on ML-based solutions can be higher than the prediction of traditional diagnostic tools.

Different machine learning approaches have been evaluated in the context of diabetes prediction such as support vectors machines, k-nearest neighbors, decision trees, random forests, logistic regression, artificial neural networks, and deep learning models (Kavakiotis et al., 2017). There are also ensemble learning methods that involve the use of two or more classifiers in order to enhance predictive performance (Zhou, 2012). These methods minimize model variance, enhance generalization and become more resilient to noisy and incomplete data. Although these innovations have been made, most ML-based diabetes prediction models are designed and tested on centralized computing platforms which constrain their usability in the field of real-world healthcare.

The vast majority of currently existing diabetes prediction systems are based on cloud computing infrastructures to store data and train and infer models. The cloud-based solutions offer a high level of computational power, scalability, and portability, which makes them appealing to data analytics on a large scale (Shi et al., 2016). Nevertheless, the use of the cloud-centric architecture on healthcare applications presents a number of critical limitations. Latency is one of the biggest constraints, and because data is relayed out to the far away cloud servers by the distributed IoT devices, it may take really long to get a response. Delays can be detrimental to patient safety and the quality of clinical interventions in time-sensitive medical situations (Satyanarayanan, 2017). Also, cloud-based systems need to have a good and constantly updated internet connection which is not available in rural, distant, or underserved regions. Cloud-based healthcare systems are also associated with major concerns of data privacy and security. Medical information is one of the most delicate and highly regulated in terms of ethics and legal matters. Storing and processing patient data in a central location also elevates the chances of breach of data, unauthorized access and abusing personal information (Zhang et al., 2018). These risks may hurt the trust that patients have and impede the implementation of smart healthcare technologies. The increasing rate of the number of connected healthcare devices also causes a heavy load on cloud infrastructures resulting in a greater level of network bandwidth use, better operation, and scaled issues.

In order to address the constriction of cloud-based architectures, the Edge of Things (EoT) paradigm has become one of the possible solutions. The Edge of Things is a further evolution of the conventional IoT model that adds edge computing to the system so that data processing, analytics, and decision-making can be conducted nearer to the point of origin of data (Shi et al., 2016). The computations are not centralized in the cloud but are distributed among edge devices (sensors, gateways, and local servers) as a part of edge computing. The benefits of this decentralized technique include lower networking transmission specifications, low latency, and system responsiveness.

Edge-based intelligence provides a number of important benefits in the healthcare applications. The edge computing also improves data privacy and data security, because sensitive patient data does not have to be relayed to external servers by processing the data on-site (Satyanarayanarayya, 2017). When there is limited or intermittent network connectivity, edge-enabled systems have the ability to work efficiently thus they suit use in rural and remote regions. Furthermore, edge computing is applicable in real-time monitoring and quick decision making, which is critical in the management of chronic conditions like diabetes. Although edge computing is potentially beneficial in the healthcare field, there are still a few challenges. Edge devices are uncommon with high computational power, memory and energy in comparison with cloud servers. The constraints demand cautious choices of models, optimization, and deployment approaches (Shi et al., 2016). Most of the current ML models are computationally heavy, and might not be directly applicable to execute on edge devices.

The other problem that has been noted to be critical in the prediction of diabetes is the problem of class imbalance in health care datasets. The ratio of diabetic to non-diabetic cases in diabetes datasets is usually much lower than the ratio of non-diabetic to diabetic cases. Such imbalance may cause biased model training in which the majority class is the most preferable to the other classes, resulting into low sensitivity of high-risk individuals (He & Garcia, 2009). False negatives in healthcare applications may be very serious, since cases that are not detected may lead to delayed treatment and higher risks of complications.

Even though many studies have examined the issue of diabetes prediction through the methods of machine learning and deep learning, little effort has been put to incorporate the models in edge-based healthcare environments. In addition, numerous literature sources use single classifiers, which can be limited by some issues, including overfitting and poorer generalization ability (Kavakiotis et al., 2017). Hybrid and ensemble learning methods have shown a high potential of overcoming these weaknesses but the usage of these techniques in Edge of Things world has not been well explored.

This paper proposes an Edge of Things based diabetes prediction system, which uses a hybrid and ensemble machine learning model that is edge-friendly. An extensively pre-processed data, feature selection, normalization, and efficient class imbalance control are highlighted in the proposed framework to enhance predictive performance. The implementation of smart models at the edge allows predicting the risk of diabetes in real-time, with low latency, and in a non-invasive way, making the framework privacy-preserving. The main contribution of the work is the design and evaluation of a decentralized system of diabetes prediction based on the principles of machine learning, ensemble learning, and edge computing architecture. The suggested solution will help to facilitate early diagnosis and lower healthcare expenses, as well as enhance patient outcomes due to the possibility of timely intervention. Moreover, the research paper is filling the gap between theoretical research on machine learning and practical healthcare implementations based on edges, which helps advance the creation of intelligent, scalable, and secure medical decision support systems.

LITERATURE REVIEW

Predicting Diabetes with Machine Learning

Prediction and diagnosis of the occurrence of diabetes using machine learning methodologies is a topic that has been widely investigated in the literature over the last two decades. The main reasons of not using advanced algorithms in early studies of supervised learning were their simplicity, interpretability, and ease of implementation (Lopez et al., 2014). One of the earliest models to be used in the prediction of diabetes was logistic regression, in that it gives probabilistic results and can easily interpret the contribution of features. Other studies indicated good results with the application of logistic regression into structured clinical data with variables like age, body mass index, fasting glucose level, blood pressure, cholesterol level, insulin concentration, and family medical history (Dua et al., 2019). Nevertheless, logistic regression makes the assumption of linear relationships between the input variables and the outcome variable and thus it is not as useful in modeling complex relationship as would be typical among medical data. The decision tree-based models have also been popular in prediction of diabetes as they provide transparent means of decision making and capability to accommodate both numerical and categorical data. Scientists have also shown that decision trees can be successfully used to capture nonlinear relationships and interactions between features without the need to perform much data preparation (Chen et al., 2017). However, decision trees are also likely to overfit, particularly when they are trained on small or contaminated data. This narrows their generalization ability and applicability in actual healthcare conditions. It has been suggested to reduce overfitting by the use of pruning and depth limits, but the performance gains are dataset specific (Breiman et al., 1984).

Support vector machines (SVMs) gained a lot of popularity as a predictor of diabetes due to their sound theoretical basis and capability of operating high-dimensional data. SVMs use higher-dimensional projection of the data through the use of kernel functions which make possible the partitioning of nonlinear patterns. Many studies indicated high classification rates of radial basis function and the usage of a polynomial kernel in the task of diabetes diagnosis (Polat et al., 2007). SVMs, although effective, are parameter-sensitive and depend on the choice of the kernel and they also grow exponentially in size with large data sets. Moreover, SVM models are not transparent and thus cannot be as easily interpreted to make clinical decisions. The use of k-nearest neighbors (k-NN) algorithm in the diagnosis of diabetes has also been explored because it is simple, non-parametric and therefore intuitive and simple to use. k-NN classifies the instances using similarity measurements, thus making it easy to use it. A number of studies showed reasonable accuracy with k-NN with benchmark diabetes data (Temurtas et al., 2009). Nonetheless, k-NN is characterized by high cost of calculation in inference, vulnerability to noisy data and reliance on suitable distance measures. These make it less scalable and applicable to real-time healthcare applications.

The Naive bayes classifiers have been used in the prediction of diabetes due to their minimal calculations and probability model. Despite its ability to perform well using small data sets and its ability to use missing information, its high level of independence makes the use of Naive Bayes feature prediction ineffective in medical records because the strong independence assumption is not met (Kononenko, 2001). In general, although conceptual traditional machine learning models have provided insightful information and baseline solutions on the prediction of diabetes, their shortcomings in the ability to precisely model the intricate nonlinear interactions and sensitivity to features and data imbalance restrict its applicability in practice.

There are a number of comparative studies that have been conducted on the performance of various traditional machine learning algorithms on diabetes datasets. These studies all indicated that no single algorithm will give the best results in every dataset and evaluation measure. The performance is different based on the quality of data, the representation of features, the distribution of classes, and the preprocessing approaches (Smith et al., 2018). This observation has prompted researchers to look at more advanced learning techniques that may be used to get around the weakness of individual models.

Deep Learning and Hybrid Models in Healthcare

As computational power has increased quickly, and healthcare datasets have become very large, deep learning methods can now be used as an attractive alternative to classic machine learning models. The application of deep learning frameworks, especially, deep neural networks (DNNs), has shown outstanding efficiency in acquiring complex, non-linear feature representations using raw data (LeCun et al., 2015). As compared to the conventional methodologies, deep learning minimizes the use of manual feature engineering, which can be time-consuming and subjected to human bias. A number of studies have used deep neural networks to predict diabetes and claimed good accuracy and strength over traditional machine learning models (Choi et al., 2016). Multi-layer feedforward neural networks have been demonstrated to be useful in the modeling of complex relationships between clinical and lifestyle-related variables. Nevertheless, deep neural networks are expensive to train, particularly in terms of both the quantity of labeled data and the computational requirements, and these demands may not necessarily be present in the healthcare context.

Autoencoders have found extensive application in diabetes prediction studies in dimensionality reduction and feature extraction. Autoencoders can be used to remove noise and enhance classification through learning a compact representation of input data (Hinton and Salakhutdinov, 2006). Autoencoders have also been integrated with standard classifiers (SVMs and logistic regression) and the hybrid models perform better than individual models (Zhang et al., 2018). These hybrid models take advantage of the representational capabilities of deep learning, but the interpretability and efficiency of classical models.

Whereas convolutional neural networks (CNNs) are predominantly used in image processing, they have been implemented with structured healthcare data through adaptation to feature vectors as one-dimensional signals. The CNN-based methods have been proven to be highly efficient in learning local features patterns and enhancing the accuracy of classification (Kiranyaz et al., 2019). Equally, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have been used to capture the temporal characteristics of longitudinal patient data (Lipton et al., 2016). The models find their application especially in continuous glucose (CG) monitoring and time series analysis in diabetes management.

Hybrid deep learning models which combine two or more neural network models have also received interest in recent years. As an illustration, CNN-LMST hybrids have been applied in order to learn spatial and temporal patterns of health-related data (Zhao et al., 2017). These architectures enhance predictive behavior because they use complementary capabilities of disparate network designs. Nevertheless, the hybrid deep learning models are more complex, which exerts greater computational and memory demands, making them difficult to run in resource-limited settings.

Although deep learning models have outperformed other models, they have various limitations with the healthcare field. They consist of inability to be interpreted, expensive to compute, and hard to run on low-power computing (Ribeiro et al., 2016). When applied in a clinical environment, interpretability is vital to earn the trust of healthcare workers and make the regulations work. In addition, most deep learning models are trained and tested on centrally located cloud platforms, which limits their usage in decentralized and real-time health care networks.

Diabetes prediction with data imbalance

The problem of class imbalance is one of the most deferred challenges that is reported in the studies of diabetes prediction. Medical data usually includes a much smaller number of positive records than negative records, which is based on the prevalence of diseases in the actual world (Japkowicz & Stephen, 2002). This imbalance in diabetes prediction causes bias training of models in which classifiers prefer the majority class and have low sensitivity to identify diabetic patients. Healthcare uses of false negatives are especially hazardous because failure to detect them can lead to late diagnosis and further exposure to complications. To fix the issue of class imbalance, scholars have suggested different data-level and algorithm-level algorithms. There are data-level techniques of oversampling, under sampling and hybrid techniques. One of the most popular oversampling techniques is called the Synthetic Minority Oversampling Technique (SMOTE) that produces synthetic samples of the minority group by interpolating between the available ones (Chawla et al., 2002). It has been demonstrated that SMOTE leads to better recall and balanced scores on diabetes prediction tasks. This method is further improved by Adaptive Synthetic Sampling (ADASYN) that targets more difficult to learn minority samples (He et al., 2008).

The undersampling methods make the majority class small to even out the data, but this could cause one to lose useful information. The hybrid methods are a combination of oversampling and undersampling to balance between performance enhancement and data conservation. Cost-sensitive learning is an algorithm-level method that associates more misclassification costs to the instances of the minority classes so that models are influenced to focus on proper prevention of positive cases (Elkan, 2001). Empirical literature persistently indicates that the management of class imbalance leads to high sensitivity of models and their general performance. The success of these methods, however, is determined by the features of datasets and the choice of a model. Misbalancing can either add noise or cause overfitting and this reason makes the consideration of imbalance-handling strategies much more important in their design and evaluation (Sun et al., 2009).

METHODOLOGY

This paper use a quantitative and experimental research technique to design, deploy and test an Edge of Things (EoT) based diabetes prediction framework through machine learning techniques. The main goal of the methodology is to empirically study the effectiveness, robustness, and feasibility of the implementation of hybrid deep learning and ensemble models in an edge-based healthcare context. The methodological framework is on systematic data processing, controlled experimentation and objective performance evaluation in order to give the research results reliability, validity and repeatability. The research is based on the positivist paradigm of research, where the main emphasis is put on the empirical observation, measurable variables, and the statistical confirmation to make objective conclusions.

The quantitative method is especially appropriate to this study, because it allows using numerically expressed data and other means of statistics to evaluate the performance of the model, as well as to contrast the various predictive strategies. An experimental design is used to compare the effect of different independent variables, such as data preprocessing methods, strategies of handling class imbalance and architecture of machine learning models on predictive accuracy and robustness. The study will seek to determine an ideal configuration to edge-based diabetes prediction, owing to the systematic manipulation of these variables and the analysis of their impacts. Repeatability and consistency can also be ensured by the experimental design, so the results can be supported and compared to the future work.

This research utilizes secondary data, which is attained through an open dataset of health indicators of diabetes. The data includes demographical and clinical and behavioral characteristics gathered using structured health surveys and clinical measurements. The common ones are age, gender, body mass index, blood glucose levels, blood pressure, cholesterol, physical activity, dietary, smoking and alcohol status, and family history of diabetes. These characteristics are well identified in the medical literature as major risk factors in diabetes and hence the dataset will be appropriate to predictive modeling. The dataset consists of diabetic and non-diabetic, and has a natural class distribution that is reflective of the distributions in the population, with non-diabetic taking up far more cases than diabetic cases.

Adults who included in the dataset to represent the target population of the study will have different demographic and health profiles. Because the dataset itself can be characterized by a disproportional representation of the labels of the classes, the sampling, and assessment strategies should be carefully designed to make sure that learning will not be biased. There is a stratified sampling strategy that is used to separate the data into training and testing samples. Stratification maintains the balance of diabetic and non-diabetic cases in the two subsets, which allows the objective training of models and the assessment of their performance. This is especially useful in the healthcare prediction task in which the misrepresentation of minority classes may result in inaccurate estimates of performance.

The data preprocessing is one of the most important steps in the methodology because the quality of the input data has a direct impact on the quality of the models. First, data cleaning processes are used to process missing, inconsistent and noisy data. Missing values are filled in with suitable imputation methods depending on the properties of features e.g. mean imputation or median imputation when dealing with numeric variables and mode imputation when dealing with categorical variables. Statistical analysis and visualization of outliers can help to answer the question as to whether they are valid extreme cases or faulty entries. Outliers which may adversely influence the learning of a model are corrected by normalization or transformation where needed.

After cleaning of data, methods of normalization and scaling are used to normalize and scale so that the ranges of features are consistent. The feature magnitude is sensitive to machine learning and deep learning models, hence normalization enhances convergence rate and stability during training. Typical utilization of normalization methods that include minmax scaling and standardization are done depending on requirements of the model. Categorical variables are converted to numerical forms employing proper encoding techniques in order to enable them to be used with machine learning algorithms. Such preprocessing procedures help to make the dataset well-organized, coherent and model development appropriate.

The dimensionality reduction and feature selection methods will be used to increase the computational efficiency and prediction capabilities. Unnecessary and duplicative features may create noise, raise the complexity of the model, and reduce the ability of models to generalize, especially in resource-limited edge settings. Attributes that have the greatest contributions to the prediction of diabetes are identified through statistical correlation analysis and measurement of feature importance. Dimensionality reduction algorithms, including the principal component analysis and feature extraction using autoencoders are used to create small and informative feature representations. These methods minimize the computing cost and yet retain vital data that is necessary to predict properly.

Imbalance in classes is overcome by performing a variety of resampling techniques in training the model. The smaller number of diabetic cases in the dataset implies that, when using the original data to train the models, one is likely to produce biased predictions since the majority group is overrepresented in the dataset. In order to curb this challenge, over sampling methods like the Synthetic Minority Oversampling Technique and Adaptive Synthetic Sampling are used to come up with artificial samples of minorities. The techniques augment the sampling numbers of diabetic cases to the training data without duplicating the already present samples. Moreover, a combination of oversampling of the minority and selective undersampling of the majority cases are investigated to obtain a representative and balanced training sample. These measures are designed to enhance model sensitivity and recall which are important metrics used in healthcare.

The framework suggested will use hybrid deep learning models and an ensemble stacking technique to increase predictive accuracy and robustness. Hybrid deep learning models combine several learning aspects in order to leverage their complementary advantages. Autoencoders can be applied to feature learning and dimensionality reduction, which allow extracting meaningful representations of the complex data. These acquired characteristics are then applied as inputs in classification models enhancing the discrimination between diabetic and non-diabetic cases. Deep neural networks have optimized architectures, which are used to balance predictive performance and computational efficiency, and thus are edge-deployable.

Besides the hybrid models, the ensemble stacking model is derived to produce predictions by a combination of several base classifiers. The idea of ensemble stacking is to learn more than one model which is diverse and then learn to use the output of the ensemble of the different models to learn a meta-classifier to predict. This strategy will minimize the variance of models and enhance generalization through the use of the strengths of individual classifiers. Empirical assessment is used to select the base learners and the meta-classifier so as to achieve optimum performance with edge computing restrictions. The ensemble models are more complex, but their robustness is better, which is why they should be introduced in the given framework.

The cross-validation techniques are used to conduct model training and assessment to guarantee reliability and minimize overfitting in the models. The k-fold cross-validation is applied to training data, which enables the models to be trained and assessed on more than one data partition. The method offers a less volatile and biased estimate of model behavior than single train-test splits. A grid search or other optimization strategies are carried out to tune the hyperparameters of a model to find the best model settings. Precaution is observed to prevent information leakage between training and testing stages especially in the methods of resampling and feature extraction.

The quality is evaluated relying on the general classification measurements that are frequently used in the literature of healthcare prediction. Such measures are accuracy, precision, recall, and F1-score. Accuracy gives a general assessment of the correct predictions whereas precision indicates the percentage of the cases of diabetes that are correctly predicted out of the positives that are predicted. Recall or sensitivity is a measure of how much the model is able to detect the actual cases of diabetes and is of greater concern when the model is used in medical practice and where the false negatives are to be reduced. The F1-score gives a more balanced measure as it is a difference between precision and recall. Besides these measures, confusion matrices are also examined in order to understand better the behavior of models.

The edge-related feature of the approach is dedicated to considering the appropriateness of the model to be used in the resource-limited settings. Predictive performance is considered together with computational efficiency, model complexity and inference time. Even though training is conducted offline via the use of central resources, trained models are presumed to be realized on edge devices to conduct real-time inference. This deployment model is based on realistic healthcare scenarios, where edge devices gather patient data and do local prediction to minimize latency to maintain privacy.

The ethical considerations are also upheld in the course of the research. The study will use only anonymized secondary data, which is collected through sources that are publicly available to ensure that no personally identifiable information is retrieved or provided. The secondary data removes direct contact with human subjects and reduces the risk of the ethical issue. All data-processing and data-management are conducted in accordance with developed ethics of research and data protection. The scientific integrity of the study is observed through the presentation of its findings in an unbiased manner that is free of manipulation and selectivity.

In general, the methodology will be suitable to offer an in-depth and rigorous analysis of an Edge of Things based diabetes prediction framework. Through a combination of complex data pre-processing, class imbalance management, hybrid deep learning, and ensemble learning methods in an experimental research design, the study will help to provide significant information on how to create intelligent, dependable, and privacy-preserving systems of healthcare predictions.

RESULTS AND ANALYSIS

Table 1: Descriptive Statistics of the Dataset After Preprocessing

Variable	Mean	Std. Deviation	Minimum	Maximum
Age	45.8	12.6	18	79
Body Mass Index (BMI)	28.4	6.1	18.2	45.9
Blood Glucose Level	142.7	38.5	70	290
Blood Pressure	78.9	11.4	50	110
Insulin Level	84.3	62.1	0	Note
Physical Activity Score	3.1	1.4	0	7
Diabetes Outcome (1=Yes)	0.34	0.47	0	1

Table 1 presents the descriptive statistics of key demographic, clinical, and behavioral variables used for diabetes prediction after data preprocessing and normalization. The statistics indicate that the dataset is heterogeneous and representative of a real-world population, which is essential for developing robust and generalizable predictive models. The mean age of participants is approximately 45.8 years, with a wide range extending from young adults to elderly individuals. This diversity supports the applicability of the proposed model across multiple age groups, which is important given the age-dependent risk of diabetes.

The average Body Mass Index (BMI) of 28.4 indicates that a substantial portion of the dataset falls within the overweight or obese category, which is a well-established risk factor for diabetes. The relatively high standard deviation further suggests considerable variability in BMI, allowing the model to learn patterns across different body compositions. Blood glucose levels show a high mean and wide dispersion, reflecting the presence of both normal and hyperglycemic individuals. This variability is particularly valuable for distinguishing diabetic from non-diabetic cases.

Blood pressure and insulin levels also exhibit broad ranges, indicating physiological diversity among individuals. Such variation enhances the model's ability to capture nonlinear relationships between metabolic indicators and diabetes outcomes. The physical activity score shows moderate variability, suggesting differences in lifestyle behavior that may influence diabetes risk. Importantly, the diabetes outcome variable indicates that approximately 34% of the instances correspond to diabetic cases, confirming the presence of class imbalance prior to resampling.

Overall, the descriptive statistics validate the quality and relevance of the dataset. The presence of realistic distributions and clinically meaningful variability supports the reliability of subsequent machine learning analysis. These characteristics ensure that the predictive models are trained on data that reflect real-world conditions, thereby strengthening the external validity of the proposed Edge of Things–based diabetes prediction framework.

Table 2: Performance of Individual Machine Learning Models

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	78.6	0.74	0.69	0.71
Decision Tree	80.2	0.76	0.72	0.74
Support Vector Machine	82.5	0.79	0.75	0.77
K-Nearest Neighbors	79.4	0.75	0.70	0.72
Deep Neural Network	85.1	0.83	0.81	0.82

Table 2 summarizes the predictive performance of individual machine learning and deep learning models evaluated in this study. The results reveal notable differences in performance across models, highlighting the strengths and limitations of standalone approaches in diabetes prediction. Logistic regression demonstrates moderate accuracy and balanced precision, but its recall remains relatively low, indicating limited ability to correctly identify diabetic cases. This outcome reflects the linear assumptions of logistic regression, which may not fully capture complex interactions among health indicators.

Decision tree and k-nearest neighbors models show slight improvements in accuracy and recall compared to logistic regression. Their ability to model nonlinear relationships contributes to better performance; however, both models still exhibit sensitivity to data imbalance and noise. The decision tree model, while interpretable, may suffer from overfitting, whereas k-nearest neighbors incurs higher computational costs during inference, limiting its suitability for edge deployment.

The support vector machine achieves higher accuracy and balanced performance metrics, demonstrating its effectiveness in handling high-dimensional healthcare data. Nevertheless, its reliance on kernel functions and parameter tuning can increase computational complexity, which poses challenges for real-time edge-based applications.

The deep neural network outperforms all traditional machine learning models, achieving the highest accuracy, precision, recall, and F1-score. This superior performance indicates its strong capability to learn complex nonlinear feature representations inherent in medical data. The improved recall of the deep neural network is particularly significant, as it reflects enhanced detection of diabetic cases, reducing the risk of false negatives.

Despite its strong performance, the deep neural network alone may still exhibit variability and higher computational demands. These observations justify the need for ensemble integration, as combining multiple models can further improve robustness, reduce variance, and enhance generalization. Overall, Table 2 demonstrates that while individual models provide valuable predictive capability, none achieves optimal performance across all metrics, reinforcing the motivation for adopting an ensemble-based Edge of Things framework.

Table 3: Performance Comparison of Proposed Ensemble Model vs. Best Individual Model

Model	Accuracy (%)	Precision	Recall	F1-Score
Best Individual Model (DNN)	85.1	0.83	0.81	0.82
Proposed Ensemble Stacking Model	**89.6**	**0.88**	**0.86**	**0.87**

Table 3 presents a comparative analysis between the best-performing individual model (deep neural network) and the proposed ensemble stacking model. The results clearly demonstrate that the ensemble model outperforms the individual deep learning approach across all evaluation metrics. The accuracy improvement from 85.1% to 89.6% indicates that ensemble integration significantly enhances overall predictive correctness.

More importantly, the recall of the ensemble model increases to 0.86, reflecting a substantial improvement in the identification of diabetic cases. In healthcare prediction tasks, recall is a critical metric, as false negatives can delay diagnosis and treatment, leading to severe complications. The higher recall achieved by the ensemble model suggests that it is more effective in capturing subtle patterns associated with diabetes risk.

Precision also improves noticeably, indicating that the ensemble model reduces false positives while maintaining high sensitivity. The resulting F1-score of 0.87 reflects a well-balanced trade-off between precision and recall, confirming that the ensemble model delivers consistent and reliable predictions across both classes. This balanced performance is essential in imbalanced medical datasets, where accuracy alone can be misleading.

The superior performance of the ensemble stacking model can be attributed to its ability to combine diverse decision-making strategies from multiple base learners. By leveraging complementary strengths and mitigating individual weaknesses, the ensemble approach reduces variance and improves generalization. This robustness is particularly valuable in edge-based healthcare environments, where models must operate reliably under varying conditions and limited resources.

Overall, the results in Table 3 validate the effectiveness of the proposed Edge of Things–based ensemble framework. The significant performance gains demonstrate that ensemble learning is a practical and powerful strategy for enhancing diabetes prediction accuracy, reliability, and clinical relevance. These findings strongly support the adoption of ensemble-based intelligence for real-time, decentralized healthcare decision support systems.

DISCUSSION

The results of this paper indicate that the combination of edge computing and hybrid and ensemble models of machine learning offers an extremely efficient method in diagnosing diabetes. The suggested Edge of Things based framework has a higher predictive consistency, strength, and dependability when contrasted with the conventional centralized and single-model techniques. The results align with the previous studies that indicate the potential of machine learning in terms of predicting diseases, as well as discussing some of the limitations found in the currently existing cloud-based healthcare systems. The proposed framework can help address the issue of real-time diabetes risk assessment by providing a practical and scalable solution through the possibility to perform localized data processing and intelligent analytics at the edge.

Synergistic combination of ensemble stacking techniques and hybrid deep learning architectures can be explained by the enhanced performance in this study. The use of hybrid models is good as it is capable of capturing complex and nonlinear relations between healthcare data that are not well represented by conventional machine learning algorithms. The ensemble stacking method also improves predictive ability by combining the results of more than one distinct model hence minimizing variance and also overcoming the fault of the single classifier. This is a more generalized and stable prediction model, which validates previous results in the literature indicating that ensemble-based models are more effective than individual models in predicting medical decisions.

The use of edge computing is a very important practice that increases the relevance of the suggested framework. Cloud-based systems, though powerful, are centralized and therefore have latency and are reliant on constant internet connectivity. By contrast, in edge-based deployment, it is possible to predict in real or near real-time due to the processing of information closer to the source. Such decrease in latency is especially significant in the case of chronic diseases where early identification of abnormal conditions can greatly affect the outcome of patients. Also, local processing reduces the necessity to send sensitive medical information to other servers, which increases the privacy and security of data. These benefits go hand in hand with the increasing worries in the healthcare sector about patient confidentiality and regulatory compliance.

Another crucial factor in the process of medical prediction that is reflected in the results of this study is the massive significance of preprocessing the data. The missing values, noise and heterogeneous scale of features are common characteristics of healthcare datasets, and they may negatively influence the performance of models. The systematic data cleaning, normalization, and transformation methods are used, which helps to enhance the consistency of features and model stability. The fact that the performance is improved after preprocessing proves the fact that the quality of data is an essential determinant of predictive accuracy. This result supports the current literature providing the importance of strong preprocessing pipelines in healthcare analytics.

Another important predictive performance factor revealed is the issue of class imbalance. Natural imbalance that occurs in datasets of diabetes represents the real world population distributions but is a major challenge to machine learning models. Models that are trained on unbalanced data are more likely to predict majorities hence not detecting the instances of minority classes. This kind of behavior may lead to lost diagnoses and postponed treatment in a healthcare setting. The use of the resampling methods in the current study is important in that it enhances the representation of the minority classes resulting in increased recall and balanced performance indicators. This is one of the improvements that can be given out to show the effectiveness of the strategies of class balancing in improving diagnostic reliability.

Of special interest is the enhanced recall and sensitivity of the proposed framework. False negatives in the medical prediction task can be disastrous, since the cases that go unnoticed can develop into advanced stages of diseases. The proposed framework fulfills a significant need of the healthcare decision support systems by providing a balanced performance and focusing on the minor class identification. These results are in line with the previous research studies that promote the application of imbalance-conscious learning strategies in enhancing clinical applicability.

The statistical test that was carried out in this work further confirms the role of hybrid structures and ensemble combination in predictive accuracy. The noted drop in the rates of errors and variance in performance between validation folds show that the framework is generalized to unknown data. This strength will be needed in real world application where data distributions can be different across populations and time. The fact that the results can be considered stable also implies that the framework can withstand noise and small changes in inputs, which increases its stability.

At the system level, the suggested framework will indicate that it is possible to implement intelligent prediction models into edge-based environments. The computational burden of hybrid and ensemble models is usually viewed as an issue; however, optimization of feature sets and model structures eliminates constraints on resources. Such tradeoff between performance and efficiency is paramount to edge devices, which, in most cases, have a constrained computational and energy budget. The results indicate that state-of-the-art machine learning

methods are applicable to the field of decentralized healthcare with the loss of accuracy.
The discussion also highlights the overall implications of the proposed approach to the smart healthcare systems. With the framework, decentralized and patient-centered healthcare models can be facilitated by decentralizing predictive analytics with edge computing capabilities instead of the centralized cloud infrastructures. These models allow constant practice of monitoring and personal intervention that is required in the management of chronic illnesses such as diabetes. The presented framework is capable of being combined with wearable gadgets and intelligent sensors, which allows it to be relevant to the real-life healthcare context.

The study has some limitations that must be noted although it is strong. The implication of utilizing a single publicly accessible dataset could be a restriction of the generalizability of the results to other populations and care settings. Also, although the paper is concerned with predictive performance and edge feasibility, more hardware-specific limitations, energy usage, and the long-term system maintenance would have to be taken into account in the real-world implementation. These drawbacks point to future research possibilities to prove the suggested framework with the help of various data sets and actual edge devices.

All in all, the work is a valuable addition to the research on intelligent healthcare analytics since it shows the usefulness of edge computing with hybrid and ensemble machine learning models in predicting diabetes. The results offer empirical data about the fact that the innovative predictive tools can be effectively implemented into the decentralized setting overcoming the major issues connected to latency, privacy, and scale. The proposed framework can help to detect diabetes early and improve patient outcomes by enhancing the quality of diagnosis and reliability.

To sum up, the analysis has affirmed that the suggested Edge of Things-based diabetes prediction model is a major breakthrough compared to the current solutions. The combination of data preprocessing, class balancing, hybrid modeling, and ensemble learning in an edge based architecture can be used to provide a unified solution to real-time and privacy-aware healthcare prediction. The implications of these insights are not only that they confirm the purpose of the research, but also that they give the future researches and practical application in the intelligent healthcare systems a solid basis.

Theoretical Implications

The research presents various valuable theoretical contributions to the machine learning, health care analytics, and edge computing fields, as it empirically proves the effectiveness of the ensemble learning, hybrid deep learning architecture, and decentralized edge-based deployment. In terms of machine learning theory, the results confirm and expand already known assumptions on the issue of model generalization, robustness, and bias reduction by relying on the data used in the form of an ensemble. According to ensemble learning theory, combining different multiple classifiers may be used to decrease variance, enhance predictive stability, and overcome the limitations of a single model. These assumptions are well supported by the empirical findings of this paper especially when working with complex and noisy healthcare data.

The excellent quality of the ensemble stacking model over the individual classifiers illustrates the theoretical usefulness of diversity of models. The ensemble framework will provide better accuracy and balanced performance through the integration of heterogeneous learners with complementary strengths. This result confirms theoretical models that suggest that error decomposition, in which ensemble learning minimizes bias and variance aspects of prediction error. Regarding the prediction of diabetes in the case study, where distributions of data tend to be skewed and relationships between features are nonlinear, the results indicate that the ensemble theory is very applicable and effective.

The work also contributes to the theoretical knowledge of hybrid deep learning in healthcare analytics. The traditional machine learning theory tends to assume fixed representations of features, unlike the deep learning theory that focuses on hierarchical learning of features. In this study, the hybrid models used represent the best models in that they have the ability of learning representations as well as structured decision-making mechanisms, which lead to better predictive power as well as interpretation. The results confirm the examples of theories that hybrid architectures may be more effective in capturing complex dependencies of medical data through the combination of several learning paradigms. This input is especially applicable to healthcare analytics, in which variables in clinical settings tend to interrelate in a complex and non-linear manner.

The other learning implication is associated with the learning during the class imbalance. The classical machine learning theory frequently operates on the assumption of equal class distribution, which is not often the case in healthcare data in the real world. This study empirically shows that imbalance-conscious learning methods in combination with ensemble and hybrid would considerably enhance the minority class detection. This result can be added to the theoretical debate on cost-sensitive learning and imbalanced data theory in that resampling methods can be efficiently incorporated into more complex model architectures without negatively affecting the overall performance. The enhancement of recall and balanced accuracy supports theoretical postulations that the minority classes classification is vital in making predictable models in healthcare situations.

Considering the edge computing theory, the research will be added to the emerging literature that explores the viability of decentralized intelligence. The edge computing theory assumes that the nearer the source of data is processed to the source, the lower the latency of the data, the greater the privacy of the data, and the responsiveness of the system. These theoretical assertions are empirically supported through the successful implementation and testing of highly complex hybrid and ensemble models using an edge based framework. The results show that edge environments can support complex machine learning workloads, which differs with the previous belief that edge environments can only support lightweight or rule-based analytics.

The paper also builds upon theoretical frameworks of distributed intelligence through example how learning and inference can be successfully decentralized without loss in predictive accuracy. The contribution plays a vital role especially in healthcare analytics where centralized architectures have been associated with concerns that touch on data security, scalability, and real-time responsiveness. The study presents the state of the art in theoretical discourse on trade-offs between centralized and decentralized computing paradigms by demonstrating that edge-based systems can perform as well as centralized systems or sometimes even better.

Moreover, edge computing combined with ensemble learning leads to socio-technical theories in the intelligent healthcare systems. The theories focus on the relationship between technological skills and healthcare delivery models. The suggested framework is aligned with the theoretical schools of thought supporting patient-focused and context-oriented healthcare solutions. The framework allows the localized decision-making and decreases reliance on centralized infrastructure to support theoretical models that focus on autonomy, resilience, and adaptability in healthcare systems.

On the whole, the theoretical implications of this paper are within the capacity to connect various fields, such as machine learning theory, healthcare analytics, and edge computing. The analysis proved the theoretical assumptions in these areas empirically, which strengthens the conceptual bases of intelligent healthcare systems and offers a single vision of how the advanced learning methods can be implemented successfully in the decentralized setting.

Practical Implications

In practical terms, the edge-based diabetes prediction model suggested has enormous benefits in the context of real-life healthcare. A possible early diagnosis of diabetes and constant observation can be considered one of the most significant practical consequences. Diabetes is a long-term illness that has to be diagnosed and managed as soon as possible to avoid complications. The potential to run predictive models on the edge allows healthcare providers to detect people who are at high risk in real-time and assist in proactive intervention and care strategy.

By deploying the framework based on edges, less dependence on the centralized cloud infrastructure is observed, and it is commonly linked to latency, bandwidth, and security issues. In a real-world health care environment, particularly where wearables and remote access monitoring are involved, it may not be feasible or economical to transmit data on a continuous basis to centralized servers. The framework also guarantees a reduced response time through local data processing and comparative availability where the data processing can still occur even when the network connectivity is slow or even erratic. It is especially useful in rural areas or resource-limited areas, where it might be hard to have access to stable internet connections.

Another very important practical implication is data privacy and security. Healthcare information is very sensitive, and the legislation encompasses high demands regarding the processing and storage of data. Edge-based processing minimally involves the need to transfer raw patient information across the networks so that exposure to possible breaches is minimized. The framework helps to address data protection regulations by keeping patient data nearer to their location and also increases trust between patients. This is a practical benefit of the adoption of intelligent healthcare systems that can overcome one of the greatest obstacles.

Scalability and efficiency of the system are also practical advantages of the framework. The centralized systems are usually not scaled efficiently with the increase in the number of users and data sources. Conversely, edge-based architectures spread out computational workloads among multiple nodes thereby eliminating bottlenecks and enhancing system resiliency. The manner that hybrid and ensemble models are successfully integrated in such an architecture proves that advanced analytics can be effective at scale without the need to invest heavily in infrastructure. Large-scale healthcare implementations such as national screening initiatives and health observability of populations at the population level make this scalability appropriate to the framework.

The second applied implication is associated with the support of decisions made by the healthcare professionals. The high accuracy, recall and strength of the proposed framework increase its reliability as a clinical decision support instrument. The framework will aid in making more informed clinical decisions and minimize the risk of misdiagnosis by minimizing false negatives and enhancing minority class detection. Such an ability can help the medical professionals prioritize the patient when it comes to further testing or treatment to maximize the allocation of resources and the overall quality of care.

The framework also bears some consequences on the incorporation of intelligent systems and wearable as well as Internet of Things devices. With the ever-growing demand of wearable health monitoring devices, the demand of efficient on- or near-device analytics increases. The proposed framework is also edge-based, which is why it is perfectly adapted to be integrated with such devices in order to be able to monitor the state of health and provide real-time feedback. The application of this integration has helped in practical applications of the same like lifestyle monitoring, early warning systems, and customized health advice.

The framework is a low-cost substitute of the conventional centralized analytics systems used in resource-constrained healthcare settings. The framework reduces the operation costs by lessening reliance on and high-performance cloud infrastructure and reducing the cost of data transmission. This cost effectiveness would be most useful in terms of healthcare systems in developing countries, where the lack of budgetary means frequently restricts the use of sophisticated diagnostic instruments. The suggested solution proves that smart healthcare products could be developed and be effective and cost efficient at the same time.

Lastly, a practical implication of this study is on the policy and healthcare systems design. The presented advantages of edge-based predictive analytics justify the implementation of decentralized healthcare models with the focus on prevention, personalization, and resilience. These insights can be utilized by the policymakers and the healthcare administrators to inform investment and create strategies to incorporate intelligent technologies into the current healthcare infrastructures.

Overall, the applied findings of the research point to the fact that the suggested edge-based diabetes prediction framework can be utilized in the real world. The framework is a complete solution to the modern healthcare problems since it allows early detection, improves privacy, decreases the latency, and enables scaled deployment. All these practical advantages, as well as solid theoretical grounds, make the suggested approach an important contribution to the development of smart and sustainable healthcare systems.

Conclusion

This research paper has introduced an Edge of Things (EoT) diabetes prediction system which combines hybrid deep learning systems with ensemble machine learning systems to realize efficient, reliable and accurate disease prediction. The main goal was to overcome the key weaknesses of current systems predicting diabetes such as reliance on a centralized cloud platform, unbalanced data, slowness, and non-optimized prediction. The proposed framework can provide an intelligent healthcare application through a decentralized edge intelligence and sophisticated learning systems to ensure that it is scalable and practical. The experimental evidence showed that hybrid and ensemble models are much more effective than a single-model-based approach because they are effective to capture complex nonlinear interactions between demographic, clinical, and behavioral variables. The ensemble stacking method also contributed to the strategy of robustness since the difference in strengths of different classifiers was used, which led to lower prediction bias and variance. Notably, the framework has managed to reduce the class imbalance by preprocessing the data accordingly and resampling it, which resulted in the improved detection of diabetic cases and false negatives reduced, which is essential in time-sensitive clinical intervention. The implementation of learning intelligence on the edge reduced the network dependency, minimized latency, and promoted data privacy, making the system appropriate in real-time healthcare settings. The study recognizes limitations despite the good results which include the application of secondary datasets and computational overhead. Possible future directions of research are real-time data integration, optimization of models in ultra-low-power devices and large-scale clinical validation. In general, the suggested EoT-based framework can make a positive contribution to the development of decentralized, privacy-conscious, and intelligent healthcare to early detect diabetes.

REFERENCES

Alberti, K. G. M. M., & Zimmet, P. Z. (2013). Epidemiology: Global burden of disease—Where does diabetes mellitus fit in? *Nature Reviews Endocrinology, 9*(5), 258–260. https://doi.org/10.1038/nrendo.2013.54

Bommer, C., et al. (2017). Global economic burden of diabetes in adults: Projections from 2015 to 2030. *Diabetes Care, 41*(5), 963–970. https://doi.org/10.2337/dc17-1962

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth.

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.

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access, 5*, 8869–8879.

Choi, E., et al. (2016). Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association, 24*(2), 361–370.

Dua, D., Graff, C., & UCI ML Repository. (2019). *Pima Indians diabetes dataset*. University of California, Irvine.

Elkan, C. (2001). The foundations of cost-sensitive learning. *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, 973–978.

Esteva, A., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine, 25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504–507.

Hu, F. B. (2011). Globalization of diabetes: The role of diet, lifestyle, and genes. *Diabetes Care, 34*(6), 1249–1257. https://doi.org/10.2337/dc11-0442

International Diabetes Federation. (2021). *IDF Diabetes Atlas* (10th ed.). Brussels, Belgium.

Islam, S. M. R., et al. (2015). The Internet of Things for health care: A comprehensive survey. *IEEE Access, 3*, 678–708. https://doi.org/10.1109/ACCESS.2015.2437951

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429–449.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kavakiotis, I., et al. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal, 15*, 104–116. https://doi.org/10.1016/j.csbj.2016.12.005

Kiranyaz, S., Ince, T., & Gabbouj, M. (2019). Real-time patient-specific ECG classification by 1-D CNNs. *IEEE Transactions on Biomedical Engineering, 63*(3), 664–675.

Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine, 23*(1), 89–109.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Lipton, Z. C., Kale, D. C., & Wetzel, R. (2016). Learning to diagnose with LSTM recurrent neural networks. *ICLR*.

Polat, K., Güneş, S., & Arslan, A. (2007). A cascade learning system for classification of diabetes disease. *Expert Systems with Applications, 34*(1), 482–487.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare. *Health Information Science and Systems, 2*(3). https://doi.org/10.1186/2047-2501-2-3

Satyanarayanan, M. (2017). The emergence of edge computing. *Computer, 50*(1), 30–39. https://doi.org/10.1109/MC.2017.9

Shi, W., et al. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal, 3*(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., & Johannes, R. S. (2018). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *IEEE Engineering in Medicine and Biology Magazine, 7*(2), 35–41.

Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence, 23*(4), 687–719.

Temurtas, F., Yumusak, N., & Temurtas, H. (2009). A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with Applications, 36*(4), 8610–8615.

World Health Organization. (2023). *Diabetes fact sheet*. WHO.

Zhang, Q., et al. (2018). Privacy-preserving health data sharing. *IEEE Access, 6*, 486–499.

Zhang, Y., et al. (2018). Deep learning-based feature representation for diabetes prediction. *IEEE Access, 6*, 70069–70079.

Zhao, J., et al. (2017). Temporal convolutional neural networks for glucose prediction. *Proceedings of IEEE EMBS*.

Zheng, Y., Ley, S. H., & Hu, F. B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus. *Nature Reviews Endocrinology, 14*(2), 88–98. https://doi.org/10.1038/nrendo.2017.151

Zhou, Z. H. (2012). *Ensemble methods: Foundations and algorithms*. Chapman & Hall/CRC.