Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 10, No. 1, March 2024, pp. 173-187
ISSN: 2338-3070, DOI: 10.26555/jiteki.v10i1.28690

An Innovative Artificial Intelligence-Based Extreme Learning Machine Based on Random Forest Classifier for Diagnosed Diabetes Mellitus

Dimas Chaerul Ekty Saputra[1], Elvaro Islami Muryadi[2], Raskmey Phann[3], Irianna Futri[4], Lismawati[5]
[1]Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand
[2]Department of Public Health, Faculty of Health Sciences, Adiwangsa University of Jambi, Jambi 36138, Indonesia
[3]Department of Mathematics and Statistics, Institute of Technology of Cambodia, Phnom Penh 120408, Cambodia
[4]Department of International Technology and Innovation Management, International College, Khon Kaen University, Khon Kaen 40002, Thailand
[5]Department Midwifery, Faculty of Health Sciences, Adiwangsa University of Jambi, Jambi 36138, Indonesia

ARTICLE INFO

Article history:

Received April 15, 2024
Revised May 01, 2024
Published May 15, 2024

Keywords:

Diabetes;
Extreme Learning Machine Forest;
Machine Learning;
Deep Learning;
Artificial Intelligence

ABSTRACT

Since 2014, the World Health Organization has accumulated data indicating that 8.5% of 18-year-olds and older have been diagnosed with diabetes. In 2019, diabetes caused the lives of 1.5 million people worldwide, with those under the age of 70 accounting for 48% of all diabetes-related deaths. It is estimated that diabetes causes an additional 460,000 deaths each year due to renal failure and that hyperglycemia contributes to about 20% of all cardiovascular disease-related deaths. Diabetes may have contributed to a 3% rise in the age-adjusted death rate between the years 2000 and 2019. In recent years, the fatality rate attributable to diabetes has increased by 13% in low- and middle-income countries. Statistics collected by the World Health Organization indicate that the number of persons diagnosed with diabetes has increased from 108 million in 1980 to 422 million in 2014. The objective of this study is to construct a model capable of diagnosing persons with diabetes reliably, correctly, and consistently. This research used secondary data offered by Kaggle. The original data came from the National Institute of Diabetes and Digestive and Kidney Diseases. Each of the up to 768 data points consists of nine characteristics and two outputs, such as diabetes and non-diabetes in the provided example. In this study, a single algorithm is constructed by integrating two separate algorithms. Random forest algorithms, which are based on machine learning, and extreme learning machines, which are based on deep learning, have generated extraordinarily accurate results. When the confusion matrix is used, 98.05% accuracy is attained. Therefore, it is feasible to conclude that the suggested method was successful in completing an adequate analysis and classifying the data.

Corresponding Author:

Dimas Chaerul Ekty Saputra, Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand
Email: dimaschaerulekty.s@kkumail.com

1. INTRODUCTION

The identification and treatment of diabetes mellitus continue to pose a significant obstacle in the field of healthcare, as its occurrence continues to increase on a global scale [1]. With the increasing prevalence of this persistent metabolic illness, there is a pressing requirement for sophisticated computational methods to assist in its timely identification and efficient treatment [2], [3], [4], [5], [6]. Considering this urgent need, scholars have progressively resorted to machine learning algorithms due to their capacity to analyze extensive quantities of data and derive significant patterns. Although there have been significant improvements in machine learning methods for detecting diabetes mellitus, there is still a noticeable deficiency in attaining both high precision and comprehensibility [7]. Conventional methods frequently face difficulties in achieving a harmonious equilibrium between these two vital elements, resulting in either excessively intricate models or restricted comprehensibility, impeding their practical usefulness in clinical environments [8], [9], [10], [11], [12], [13]. Furthermore, the dependence on individual classifiers may not comprehensively encompass the intricacy and diversity of diabetes mellitus datasets, hence constraining their ability to differentiate between different categories [14].

Diabetes mellitus, sometimes known as diabetes, is a disease or a chronic metabolic condition characterized by abnormalities in glucose, lipid, and protein metabolism resulting from inadequate insulin activity [15]. Diabetes mellitus (DM) may be a condition or a disease. The most notable sign of diabetes mellitus is elevated blood sugar levels [16]. Insulin dysfunction may result from a disruption or deficit in insulin production by the pancreatic Langerhans beta cells, or from a lack of insulin responsiveness in the body's cells [17]. The pancreas generates the hormone insulin, which controls the blood sugar level (glucose) [18]. In contrast, diabetics have pancreases that are incapable of secreting insulin in response to the body's needs [19]. Without insulin, cells cannot absorb glucose and convert it to energy [20]. When glucose is inadequately absorbed by cells, it accumulates in the blood [21]. This disease may result in several organ dysfunctions [21]. Inadequate management of diabetes may result in life-threatening complications [21]. Access to affordable medicine, especially insulin, is essential for persons with diabetes. A strategy to lower the number of people suffering from diabetes and obesity by 2025 has won worldwide agreement [22]. Diabetes affects around 422 million people worldwide, the great majority of whom live in low- or middle-income nations. It is estimated that approximately 1.5 million individuals die annually as a direct consequence of diabetes [22]. In the last several decades, there has been a rise in both the number of diabetes cases that have been identified and the disease's overall prevalence [23].

According to the International Diabetes Federation (IDF), by 2045 there will be 28.57 million diabetics in Indonesia [24]. This number is 47% more than the 19.47 million projected for 2021 [25]. The number of diabetics in 2021 has increased significantly during the last decade. Compared to the amount of diabetes in 2011, which was 7.29 million, the number of diabetics has climbed by 167%, to 14.14 million [26]. This increase is much more than the growth between 2000 and 2011 [26]. From 5.65 million in 2000, the number of diabetes increased by 29% throughout this time period. By 2021, diabetes will be the cause of 236,711 deaths in Indonesia [25]. The number increased by 58% from 149,871 in 2011 [26]. The IDF projects that by 2045, the worldwide diabetes population would reach 783,7 million people [27]. This is a 46% increase over the previous year's total of 536.6 million [27].

Diabetes mellitus type 1 and diabetes mellitus type 2 are two distinct kinds of diabetes. The development of type 1 diabetes is a natural result of this process since the pancreatic cells that make insulin are the target of the immune system [28]. This results in a rise in blood glucose levels, which causes organ damage. The most widespread kind of diabetes is type 1 diabetes, often known as autoimmune diabetes. The specific causes of type 1 diabetes are not yet completely known. Nevertheless, there is a theory that both hereditary and environmental factors contribute to this disease [28]. The great majority of diabetics are diagnosed with type 2 diabetes, which accounts for 90–95% of all diabetes cases [29]. Because it makes the body's own insulin production ineffective, the development of insulin resistance in cells may lead to type 2 diabetes. Alternately, this condition might be referred to as insulin resistance. Two of the methods for preventing diabetes include living a healthy lifestyle and consuming a well-balanced diet. Diabetes testing is uncomplicated. During this examination, blood sugar levels will be determined [30]. To preserve their health and quality of life, prediabetic patients must be recognized and started treatment as soon as feasible. In addition to laboratory testing, an approach that relies on disease symptoms will significantly improve both the diagnosis and prevention of diabetes.

Using current and historical data, predictive analytics combines machine learning algorithms and statistical methodologies to find information and forecast future occurrences [31]. The use of predictive analysis to healthcare-related data provides critical assessments and forecasts. Machine learning and regression methods may be used to do predictive analytics. Our research aims to tackle these issues by introducing a new method: an Extreme Learning Machine (ELM) that utilizes a Random Forest (RF) for diagnosing diabetes mellitus. The primary objective of this novel framework is to address the current disparity by capitalizing on the respective advantages of ELM and RF algorithms, consequently augmenting the precision and comprehensibility of diabetes diagnosis. Our proposed methodology combines the enhanced generalization capacity of ELM with the ensemble learning technique of RFC. This integration results in a resilient and effective solution for reliably detecting individuals with diabetes, while also offering valuable insights into the decision-making process.

The main significance of this study resides in its innovative integration of ELM and RFC, which leverages the synergistic capabilities of these algorithms to enhance diagnostic precision while maintaining interpretability. In addition, our methodology effectively tackles the constraints of current techniques by integrating a wide array of characteristics and adeptly managing the intricacies inherent in diabetes mellitus datasets. The effectiveness and superiority of our suggested strategy in detecting diabetes mellitus are demonstrated through rigorous experimentation and validation using real-world clinical data, surpassing existing state-of-the-art methodologies. The following is a list of the article's most significant contributions: (a) Construct a model that is straightforward to comprehend; (b) determine which attributes are most important to improve interpretability; and (c) use a variety of machine learning and deep learning algorithms to build an ensemble model that combines Machine Learning and Deep Learning.

The current document is organized into five separate sections. In the first section, an introductory segment is presented, which offers a comprehensive overview of the background of the subject, describes the aims, and emphasizes the research contributions. Section 2 entails a thorough analysis of the current body of knowledge. Section 3 presents a thorough exposition of the materials and methodologies utilized in the research. In Section 4, the research findings are presented, and a thorough analysis and interpretation of these data is conducted. The conclusions derived from the findings and suggestions for future research endeavors are presented in Section 5 of this study.

2. LITERATURE REVIEW

The identification of diabetes has received considerable focus in the research community, especially in the context of e-healthcare systems [32]. Conventional diagnostic systems encounter difficulties such as lengthy computational processes and limited accuracy in predictions [33]. Machine learning techniques have been identified as a possible approach to improve diagnostic capabilities in healthcare services [34]. To tackle these difficulties, the authors put up an innovative diagnostic approach that utilizes machine learning techniques specifically designed for the identification of diabetes [35], [36]. The suggested system was thoroughly evaluated using different model validation techniques, such as hold out, K-fold, and leave one subject out methods, on a clinical dataset obtained from patient histories [37]. The system's efficacy was evaluated using performance indicators such as accuracy, specificity, sensitivity, F1-score, receiver operating characteristic curve, and execution time [37], [38], [39]. A novel approach for feature selection was developed, utilizing the Decision Tree algorithm, and augmented with ensemble learning techniques such as Ada Boost and Random Forest. The proposed methodology was shown to be effective, as demonstrated by a comparative analysis using wrapper-based feature selection algorithms. The experimental results showed enhanced classification ability, which was attributed to the selection of crucial characteristics, particularly Plasma glucose concentrations, Diabetes pedigree function, and Body mass index. Crucially, the suggested approach demonstrated better performance in comparison to current cutting-edge techniques, confirming its potential for use in e-healthcare settings. The statistical studies provided additional evidence supporting the system's ability to properly detect diabetes, therefore emphasizing its potential as a valuable tool in clinical practice.

Diabetes is a metabolic disorder that may be caused by a lack of insulin synthesis or an inability to properly react to insulin. Insulin is a hormone that plays a crucial function in controlling blood sugar levels. This condition is characterized by symptoms such as polyuria, polydipsia, rapid weight loss, impaired vision, and fatigue [40]. This leads to severe health issues, such as abortions, blindness, strokes, and organ failure. Recommendations from the medical community encourage early diagnosis to identify at-risk individuals and proactive lifestyle monitoring by patients to reduce risk factors. This action is performed as part of an attempt to reduce the number of potential problems. Traditional ways of controlling chronic illnesses entail using rule engines or score estimates to determine whether a patient is at risk for getting diabetes. These techniques are far less successful than machine learning algorithms for detecting a patient's potential health risks [41]. Recent advances in machine learning and feature engineering have created new opportunities for improving the early diagnosis and treatment results for patients with chronic diabetes [33]. A recent study on the treatment of diabetes has shown that the most popular system types include insulin self-management software, automated SMS risk alerts, wearable blood glucose monitors, and virtual diagnostician coaching [42]. Ramesh [43] utilized PIMA Indian Diabetes Dataset includes 768 data. The researcher used four machine learning techniques, including KNN, LR, Gaussian NB, and SVM. The major purpose of this study is to collect patient information utilizing data gathered from personal healthcare equipment and consumer wearable devices. This would make it feasible to detect diabetes risk early and keep physicians informed of the most current developments. With an accuracy rating of 83.20%, the SVM RBF kernel produced the most accurate results. Hassan [44] with 250 data and 16 unique attributes has given good results. This study employs three distinct machine learning models, namely LR, SVM, and random tree. The findings of ten-fold cross-validation revealed that random tree achieved the highest overall accuracy, 97.50% [44]. Sivaranjani [45] the likelihood of developing diabetes-related disorders was evaluated by the use of the machine learning approaches of support vector machine (SVM), as well as random forest (RF). The principal component analysis (PCA) technique will be used to lower the total number of dimensions after the significance of various factors has been established. RF reaches an accuracy of 83% while processing data from the diabetes database maintained by PIMA [45]. Khanam [46] Utilized information from the UCI Machine Learning Repository-accessible PIDD data collection. The data collection contains details on 768 patients and the nine identifying traits they all share. They utilized the data to train seven machine learning algorithms to generate diabetes-related predictions. The researcher uses seven methods, such as DT, RF, NB, KNN, LR, NN, and SVM. The best accuracy is NN with 2 hidden layers which produced an accuracy of 88.6% [46].

In the field of healthcare, it is crucial to identify diseases at an early stage, especially when it comes to the hidden danger of diabetes. This condition is a worldwide health issue that is worsened by modern eating habits that are high in sugar and fat. In order to address this issue, a state-of-the-art method utilising machine learning (ML) algorithms has emerged as a promising solution for precise disease prediction. This paper introduces a novel machine learning approach designed exclusively for predicting diabetes, which represents a significant change in how predictive analytics is applied in the medical field. This new framework integrates two distinct models, namely Support Vector Machine (SVM) and Artificial Neural Network (ANN). These models carefully analyse complex statistics to determine the probability of a positive or negative diabetes diagnosis. The dataset is carefully divided into separate training and testing subsets, guaranteeing thorough model training and rigorous evaluation. The SVM and ANN models produce outputs that are used as input membership functions for a complex fuzzy logic-based model. This model serves as the ultimate decision-maker in deciding the diagnosis of diabetes. The fused model is cleverly and easily incorporated into a cloud storage system, marking the beginning of a new era of easy access and effectiveness in predictive healthcare analytics. This allows for immediate forecasts using patients' medical records, giving clinicians the ability to take early action based on valuable information. The suggested fused ML model has a remarkable prediction accuracy of 94.87%, surpassing the effectiveness of previously published approaches. This innovative methodology challenges traditional disease prediction methods, highlighting the revolutionary power of machine learning in healthcare. It also signals a promising new era in proactive disease treatment and prevention efforts.

3. METHODS

This section will provide a clear explanation of the methods used to collect data, outline the research process, and discuss the use of different machine learning algorithms, including Support Vector Machine, Random Forest, K-Nearest Neighbor, and Extreme Learning Machine. Moreover, provides the suggested Extreme Learning Machine Forest model, highlighting its specific use in diabetes research. To establish scientific legitimacy and acceptance, it is essential to support the narrative with suitable references.

3.1. Data Collection

The Pima Indian Diabetes Dataset provided on Kaggle was used for the objectives of this study. This research utilizes 768 data with nine criteria. The data set is separated into two unique groups: normal, with 500 entries, and diabetes, with 268 records. The training data represent 80% of the total, whereas the test data represent 20%. The National Institute of Diabetes and Digestive and Kidney Diseases was responsible for the creation of this content. The objective of this data collection is to identify whether a patient has diabetes by using the several diagnostic tests available within the data set. The selection of these specific samples from a larger database was controlled by several characteristics and limits. To enroll in the therapy program, patients must be at least 21 years old and of Pima Indian descent. The Pima Indian Diabetes Dataset comprises information on 768 distinct Phoenix, Arizona-native women. 258 people's diabetes tests returned positive findings, whereas 500 people's tests returned negative results. One dependent variable (goal) is accompanied by eight features: pregnancies, the Oral Glucose Tolerance Test (OGTT), blood pressure, skin thickness, insulin, Body Mass Index (BMI), age, and pedigree diabetes function. The objective is to identify whether diabetes runs in a family. Once every two years since 1965, the National Institute of Diabetes and Digestive and Kidney Diseases has conducted a demographic study among the Pima people. The Pima Indians Diabetes Dataset contains information about traits that may or may not be associated with the onset of diabetes and its potential consequences. Genetic and environmental factors combine to create type 2 diabetes, according to an epidemiological study. This information is provided by the Diabetes Dataset for Pima Indians. As the recommended model, this study utilizes several direct Machine Learning and Deep Learning approaches, including the support vector machine (SVM), the random forest (RF), the k-nearest neighbor (KNN), the extreme learning machine (ELM), and the extreme learning machine forest (ELMf).

3.2. Research Flow

This study utilizes 768 data with nine criteria. The database-entered data will subsequently undergo data preprocessing, which consists of a MinMax scaler, one-hot encoded, input missing data value, and LabelEncoder. After the data has been preprocessed, it is divided into its two component parts: the training data, which accounts for 80% of the total, and the test data, which accounts for 20%. The training data will be examined using a range of machine learning algorithms, including deep learning, and ensemble machine learning with deep learning techniques. After evaluating the training data, the accuracy of each approach will be determined. The most precise technique will be suggested as the paradigm for this investigation. The provided test data is next evaluated using the model with the best accuracy offered. After the testing procedure has been completed, the suitable class is retested to evaluate the performance outcomes of the used approach. Once performance results are obtained, the procedure is complete. In Fig. 1 explain the diagram of this research flow.

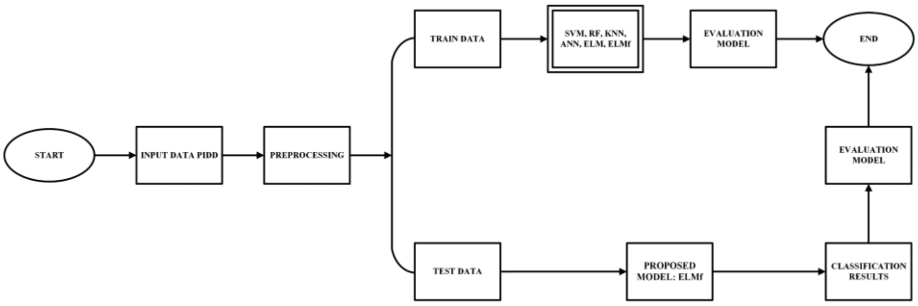

Fig. 1. Research Flow Diagram

3.3. Support Vector Machine

The support vector machine, abbreviated as SVM for simplicity's sake, is now one of the most widely used algorithms for supervised learning classification and regression [47]. Machine learning classifies most issues. The Support Vector Machine (SVM) algorithm classifies data points by constructing the ideal line or decision boundary that separates n-dimensional space into classes [47]. This limit is denoted by a hyperplane as the best possible alternative. SVM can identify both the hyperplane's extremities and the vectors that contribute to its formation. These extreme occurrences are known as support vectors, and the technology associated with them is referred to as a support vector machine [48]. Given a set of labeled training data (x_i, y_i), where x_i is the i-th data point in the feature space and y_i is its corresponding class label ($y_i \in \{-1, +1\}$), the goal of SVM is to find the optimal hyperplane $w \cdot x + b = 0$ that maximizes the margin between the classes while minimizing the classification error. The decision function for a new data point x is given by:

$$F(x) = sign(w \cdot x + b) \tag{1}$$

where w is the weight vector perpendicular to the hyperplane, b is the bias term, and $sign(\cdot)$ is the function, which assigns $+1$ is the input is positive, -1 is it's negative, and 0 if it's zero. The distance from a point x to the hyperplane is given by:

$$X(x, w, b) = \frac{|w \cdot x + b|}{||w||} \tag{2}$$

where $||w||$ is the Euclidean norm of the weight vector w.

The margin is the distance from the hyperplane to the closest data point of any class, and SVM aims to maximize this margin, the margin given by:

$$M = \frac{2}{||w||} \tag{3}$$

The optimization problem of SVM can be formulated as finding w and b that minimize $||w||$ subject to the constraint that for each data point (x_i, y_i):

$$y_i(w \cdot x_i + b) \geq 1 \tag{4}$$

This constraint ensures that all data points are correctly classified and are on the correct side of the decisions boundary with a margin of at least 1. This is referred to as the “hard margin” SVM.

3.4. Random Forest

Random forests are used to tackle regression and classification issues in the machine learning [49]. It does this via the use of ensemble learning, a technique that handles complex issues by combining the efforts of numerous separate classifiers [50]. A random forest strategy is comprised of several decision trees. Bagging or bootstrap aggregation may be used to train the "forest" component of the random forest method [51]. The ensemble approach of meta-algorithm bagging may increase the accuracy of machine learning algorithm. Based on evaluations of decision trees, the algorithm (random forest) decides. Calculate and average the average yield of several trees [51]. Increasing the number of trees in the forest may increase the result's precision. The random forest technique avoids the disadvantages associated with the decision tree method. It improves precision and reduces the amount of dataset overfitting [51].

3.5. K-Nearest Neighbor

K-Nearest Neighbors (KNN) is a kind of supervised, non-parametric learning classifier that uses the distance between two data points to predict or categorize the grouping of one data point [52]. It may be used to handle regression problems as well as classification problems; however, it is often used as a classification technique since it is based on the notion that comparable locations are frequently situated adjacent to one another [52]. The KNN algorithm predicts the value of a new data point by comparing its similarity to the neighboring data points in the feature space. During the training process, the KNN algorithm effectively retains all the available data points together with their respective class labels or goal values. When predicting a new data point, the algorithm computes the distance between the new point and every other point in the training set, usually using metrics such as Euclidean distance. The k nearest neighbors to the new data point are selected based on these distances. In classification tasks, the predicted class label is established through a process of majority voting among the k nearest neighbors. In contrast, for regression tasks, the predicted value is calculated as the average of the target values of those neighbors. The pivotal hyperparameter in KNN is k, which denotes the quantity of neighboring data points to consider. Selecting an optimal value for k is crucial for the model's performance and capacity to generalize. The simplicity and intuitiveness of KNN make it a widely favored option, particularly in situations when the decision boundary is intricate, or the data distribution is erratic. Nevertheless, the performance of the system can be influenced by the selection of k and the distance metric, and it may not handle huge datasets efficiently due to its computational complexity.

In mathematical terms, let D be the training dataset with N data points, and x be the new data point for which we want to make predictions. The prediction $\hat{y}$ with KNN can be defined as follows

$$\hat{y} = argmax_y \sum_{i=1}^{k} \delta(y, y_i) \tag{5}$$

where δ is the Kronecker delta function, y_i is the class label of i-th nearest neighbor, and k is the number of nearest neighbors.

Moreover, the performance of KNN can be influenced by the selection of a distance measure and the use of data pretreatment procedures, such as normalization or standardization. Although KNN is a simple algorithm, it can be remarkably efficient, particularly in situations when the decision border is irregular, or the data distribution is ambiguous.

3.6. Extreme Learning Machine

The Extreme Learning Machine (ELM) is frequently used in batch learning, sequential learning, and incremental learning because to its rapid and efficient learning speed, rapid convergence, great generalization capabilities, and simplicity of implementation. This is a result of the ELM's rapid and effective learning rate [53]. ELM is an abbreviation for a model based on the multilayer perceptron's topological architecture (MLP) [53]. When compared to the MLP, the ELM is characterized by the random generation of its weights and biases. As a direct result of the creation of the original ELM, the use of ELM has expanded to cover semi-supervised learning and unsupervised learning in addition to supervised learning [54], [55]. Due to the memory residency of the conventional ELM as well as its high spatial and temporal complexity, it is not viable to train vast amounts of data in a fast and effective way [56].

3.7. The Proposed Model: Extreme Learning Machine Forest

The newly proposed ELMf (extreme learning machine forest) model consists of two distinct components. While the forest part functions as a feature detector to identify data pattern representations of raw inputs by observing training results, the ELM section functions as a learner to predict outcomes by using newly found feature representations. These duties are carried out by the network as a whole. Both of these components are essential to the model's overall operation. During the forest phase, self-sufficient decision trees are created, changing the forest into a forest of trees. The Random Forest model is hence the natural choice when attempting to produce a forest. In addition, a number of distinct forest architectures are feasible. For instance, if the feature space is sorted and known, one may use a network-guided forest or generate a forest quickly by bagging trees. If the feature space is known, these two choices are accessible [34].

Random forests are the only kind of feature detector used in this study. The ELMf classifier's training is divided into two parts. In the first stage, labeled training data are used to match the forest, and in the second step, predictions from each tree in the forest are used to train the fully connected ELM. In the initial stage, the forest is matched with labeled training data. Matching the forest with the aid of labeled training data is the first step. After undergoing a two-stage training procedure, the installed forest and ELM will provide test predictions using the test sample for the whole model when given a test sample. The ELMf model views the forest F as a collection of decision trees organized in the following manner:

$$F(\theta) = \{J_m(\theta_m)\}, \qquad m = 1,2,3,\dots,M, \tag{6}$$

where M is the total number of trees in the forest, $\theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_M\}$ reflects the parameters inside F within random forest, θ involves dividing variables and their corresponding values. In the course of identifying characteristics, M conforms to the training data X and y, where $X \in R^{n\times p}$, n samples and p characteristics comprise the input data matrix. $y \in R^n$ is the result vector containing the label for categorization. Through the appropriate woodland for observation x_i, In (6), we extract the forecast from each tree in F, then:

$$f(x_i;\theta) = (T_1(x_i;\theta_1), \dots, T_M(x_i;\theta_M))^T, \tag{7}$$

where $T_M(x_i;\theta_M) = \hat{y}_{iM}$ the binary forecast of observation x_i supplied by J_m. f_i is a binary vector that summarizes the forest signal and subsequently acts as the ELM's new input characteristics. o_j is the result obtained by (8):

$$o_j = \sum_{j=1}^{N} \beta_j g\left(\sum_{j=1}^{N} w_j x_i + b_j\right), \tag{8}$$

where the following additional feature representations are supplied by the forest in (7), besides x_i represents ith, $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$ indicates the combined weight of the hidden and output layers, $w_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$ reflects the difference in weight between the jth hidden layer and the ith input layer, b_j denotes the jth hidden layer's threshold and $g(.)$ represents the activation function using the sigmoid function. Accordingly, the output matrix of hidden layers H Utilizing the sigmoid function, depicts the activation function. Accordingly, the output matrix of hidden layers H and output-hidden layer weights b for the specified input-output sample pairs is now computed as follows $H\beta = O$. H for the transfer function $g(.)$ as in (8).

The $w_j . x_i$ is portion becomes the sum of the components of another.w_j and x_i. The N (8) may be expressed concisely as:

$$H\beta = O, \tag{9}$$

where,

$$H = \begin{bmatrix} g(w_1 x_1 + b_1) & \cdots & g(w_N x_1 + b_N) \\ \vdots & \ddots & \vdots \\ g(w_1 x_n + b_1) & \cdots & g(w_N x_n + b_N) \end{bmatrix}_{n\times N},$$
$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{N\times m}, \tag{10}$$

$$O = \begin{bmatrix} o_1^T \\ \vdots \\ o_N^T \end{bmatrix}_{N \times m} .$$

The H represents the output matrix of the hidden layer. H^T inverse of H according to the generalized Moore–Penrose formula and t is the target class/data label in (11),

$$\hat{\beta} = H^T t. \tag{11}$$

Nevertheless, the output weights are calculated via a mathematical transformation, resulting in a decrease in the amount of time required in the training phase, during which the network's parameters are modified periodically using learning parameters (such as learning rate and iteration). The ELMf model's architecture is seen in Fig. 2.

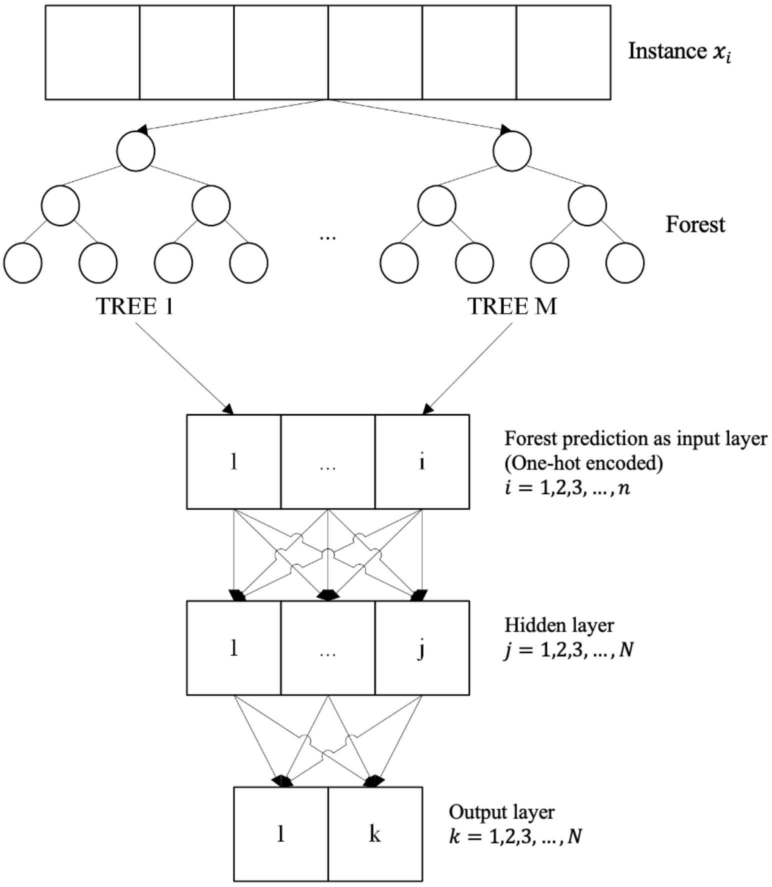

Fig. 2. ELMf Architecture

3.8. Diabetes

Diabetes is a chronic disorder that expresses itself in one of two ways: either the pancreas cannot generate an appropriate quantity of insulin, or the body cannot use the insulin it does produce [40]. Insulin is a hormone that regulates the amount of glucose circulating in the body [57]. Hyperglycemia, another term for high blood glucose or blood sugar, is virtually often a consequence of uncontrolled diabetes. Hyperglycemia causes extensive damage to several physiological systems, including neurons and blood vessels, over time [58]. Diabetes was diagnosed in 8.5% of persons aged 18 and older in 2014 [59]. In 2019, 1.5 million deaths were directly linked to diabetes, with 48 percent of all diabetes-related deaths happening in those younger than 70 years old [59].

An additional 460,000 renal disease-related deaths were attributed to diabetes, and increased blood glucose leads to about 20% of cardiovascular mortality [16]. The age-standardized mortality rate attributable to diabetes rose by 3% between 2000 and 2019 [38], [39]. In countries with a lower middle income, diabetes-related mortality increased by 13%. Between 2000 and 2019, the risk of dying between the ages of 30 and 70 from any of the four most prevalent noncommunicable diseases (cardiovascular diseases, cancer, chronic respiratory disorders, and diabetes) decreased by 22%. These include cardiovascular problems, cancer, and chronic respiratory conditions [39]. The chronic condition known as diabetes is characterized by blood glucose levels that are too high [40]. There are two fundamental types of diabetes:

1. Insufficient insulin production characterizes type 1 diabetes, which needs daily insulin injections. In the past, diabetes mellitus, insulin-dependent diabetes, and infantile diabetes have all been used to refer to this illness. In 2017, 9 million people were diagnosed with type 1 diabetes; the great majority of these individuals reside in high-income countries. It is uncertain what causes it and what preventative measures should be performed. Among the symptoms include polyuria, polydipsia, persistent hunger, weight loss, eyesight issues, and fatigue. It is conceivable that these symptoms will manifest spontaneously [62].
2. Insulin resistance is one of the leading causes of type 2 diabetes (previously known as non-insulin-dependent or adult-onset diabetes). In more than ninety-five percent of patients, type 2 diabetes is diagnosed. This kind of diabetes is often caused by a combination of a high body fat percentage and insufficient physical activity. There is a chance that symptoms may resemble those of type 1 diabetes, although they are often milder. As a result, the ailment could not be diagnosed until many years after it first emerged, by which time it had already caused severe harm. In the past, only adults were diagnosed with this kind of diabetes, but in recent years, a considerable rise in instances has been seen among younger patients as well [29].

The occurrence of diabetes type 2 is seen in a much greater number of people than in diabetes type 1 [16]. Some pregnant women have blood glucose levels so high that their bodies are unable to produce sufficient insulin to absorb it all [63].

4. RESULTS AND DISCUSSION

In this paper, an experimental investigation was conducted to identify and categorize illness in a dataset of diabetic patients using the suggested ELMf model. During the experimental work, the real diabetes dataset was used and separated into training and testing sets. Utilizing a Python-written Confusion Matrix, the classification method in the research was evaluated. In the research, the implementation was performed on an Apple M1 machine with 512GB of internal memory and 8GB of RAM.

4.1. Evaluation Model

We separated the data into two groups throughout the testing process: test data and training data. Evaluation is conducted by applying the Confusion Matrix model to both the training data and the test data. When analyzing the accuracy of classification systems, a confusion matrix is a valuable tool that is often used. This research employs a data collection that is divided into five unique groups, including those with and without diabetes. 80% of the data used for training and 20% of the data used for testing were included in the dataset for this study. Comparing the efficacy of various classification algorithms may be accomplished by examining their Accuracy, Precision, Sensitivity, and F1-Score values (9)–(12). The method of value distribution is broken out in Table 1 of the Confusion Matrix, which is located here:

Table 1. Confusion Matrix

Real Class **Prediction**	**True**	**False**
True	True Positive	False Negative
False	False Positve	True Negative

It can produce scores for precision, recall, and F1-Score depending on the characteristics listed above. Value-calculating formulas may be characterized as follows:

$$\text{Accuracy} : \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$\text{Precision} : \frac{TP}{TP + FP} \tag{13}$$

$$\text{Sensitivity} : \frac{TP}{TP + FN} \tag{14}$$

$$\text{F1-Score} \quad : \quad 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{15}$$

4.2. Experimental Results of Extreme Learning Machine Forest

In this article, the ELMf classification model is suggested as a feasible alternative for the diabetes dataset. Combining the Extreme Learning Machine with Random Forest algorithms yields the ELMf model. Hidden layer, the ELM model has a single feedforward network, while the random forest model use entropy as a selection criterion. Using accuracy, sensitivity, precision, and the F1-score, their utility for identifying diabetes datasets was evaluated. Table 2 presents the ELMf model's results. In the two-class diabetes dataset, the ELMf model with 80% training data and 20% test data, 98.05% accuracy, 97.99% precision, 97.40% sensitivity, and 97.68% F1-Score produced the most accurate results.

Table 2. Performance Results of ELMf Model

Model	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
ELMf	98.05	97.40	97.99	97.68

Table 3 provides the performance index of each class as well as the suggested method with the best success rate on the diabetes dataset. Diabetes classes had the greatest prediction accuracy in the random forest, k-nearest neighbor, support vector machine, extreme learning machine, and extreme learning machine forest algorithms. For PIDD, the ELMf algorithm is effective.

These data are a perfect fit for the ELMf algorithm. Multiple parameters, such as the results of the performance assessment of the machine learning technique employed prior to its conjunction with the ELM algorithm for deep learning, are considered throughout the merging process. These variables affect the final assessment outcomes of a combined technique, and this research reveals that the ELMf algorithm works very well on PIDD.

Table 3. Index of performance results for each class in each method

Model	Classes	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
SVM	Diabetes	76.62	62.22	59.57	60.87
	Non-Diabetes		85.57	84.11	83.33
RF	Diabetes	81.82	65.96	72.09	68.89
	Non-Diabetes		88.78	85.58	87.15
KNN	Diabetes	79.22	59.57	68.29	63.63
	Non-Diabetes		87.85	83.18	85.45
ELM	Diabetes	95.45	93.44	95	94.21
	Non-Diabetes		96.77	95.74	96.25
ELMf	**Diabetes**	**98.05**	**95.74**	**97.83**	**96.77**
	Non-Diabetes		**99.06**	**98.15**	**98.60**

In the medical field, misdiagnosing healthy patients as unwell and vice versa is very dangerous. In this regard, the use of data mining technologies for accurate diagnosis is increasing. This study aims to correctly classify and diagnose diabetes using an extreme learning machine forest model, as well as to develop a decision support system to assist doctors' decisions. A condition known as diabetes, sometimes known as diabetes mellitus, is characterized by an unusually high level of glucose in the blood. Either the body is unable to make any insulin, is unable to produce a considerable quantity of insulin, or is unable to effectively use the insulin that it does produce. All three of these issues are diabetic symptoms. Over time, high glucose levels may damage blood vessels and neurons, leading to cardiovascular, renal, ophthalmic, and foot problems. Diabetes is predicted to impact 415 million people globally, or one out of every eleven adults. 46% of individuals with diabetes are undiagnosed. Diabetes is five times more frequent among Pakistani women, The frequency is two-and-a-half times more among Indian women and four times greater among Bangladeshi and Indian persons overall. South Asians have a prevalence of diabetes that is often six times higher than Africans, who have a prevalence of diabetes that is typically three times higher. It is uncertain why this is the case, however, various risk factors may be responsible. Poor economic situations and a sedentary lifestyle have been related to unhealthy eating habits; genetics may also play a role. Consequently, several scientists have developed artificial intelligence for diabetes prediction.

Diabetes research has made substantial use of machine learning and deep learning, using a range of different techniques. In the plethora of diabetes datasets that may be acquired from the plethora of publicly available sources, varying kinds and quantities of data can be discovered. This is shown by the PIMA Diabetes Dataset (PPID). The Phoenix Pregnancy Incidence Database (PPID), which was first produced by the National Institute of Diabetes and Digestive and Kidney Diseases, has information on 768 Arizona women who lived in the Phoenix area. The PPID dataset has been widely used in several studies using various approaches. The computer method can anticipate diabetes in its earliest stages. Throughout medical history, several algorithm-based techniques based on machine learning have been developed to facilitate the rapid and accurate identification of disease. There have been developed several research approaches, like as SVM [48], ANN [64], AdaBoost [44], RF [45], [51], Bagging [65], and DNN [66], [67], [68]. Table 4 shows the competitive results using PIDD compared with another model.

Table 4. Competitive results from another method

Authors	Year	Dataset	Method	Accuracy (%)
Joshi *et al.* [69]	2021	PIDD	**LR** and ML	78.26
Gupta *et al.* [70]	2022	PIDD	QML and **DL**	95.00
Howlader *et al.* [71]	2022	PIDD	GAMBoost, RLR, PMR, BGLM, PLR, GLM, SDWD, **GBR**, GAMLOESS, NB	90.91
Kibria *et al.* [72]	2022	PIDD	ANN, RF, SVM, LR, AdaBoost, XGBoost, and **Ensemble (RF, XGB)**	90.00
Saputra *et al.* [38]	2023	PIDD	SVM, RF, KNN, **SMKSVM-RF**	73.37
Proposed Model	**20 24**	**PIDD**	**Extreme Learning Machine Forest**	**98.05**

5. CONCLUSION

Diabetes is a widely widespread medical disorder that impacts a substantial population. The development of diabetes in people may be influenced by two factors: family history and lifestyle choices. The distinction between those with diabetes and those without diabetes remains a subject of debate that is both captivating and contentious. The prevalence of risk factors associated with the development of diabetes is on the rise. In the context of this study, the use of artificial intelligence was determined to be advantageous in acquiring knowledge of patterns and delivering a notable degree of precision. By employing machine learning and deep learning techniques, it has been proven that it is feasible to merge two separate learning methods to create a third strategy that is both specialized and efficient. Concerning the progression of knowledge, this research has made several notable contributions. The employment of a new technology called ELMf has the potential to enhance the accuracy of a confusion matrix to 98.05%. We are certain that this approach will serve as a valuable supplement to the existing ways, and that the formula now in use has the potential to become the most optimal healthcare system currently accessible.

Acknowledgments

The research conducted in this study received partial financial assistance from the Khon Kaen University Grant for ASEAN Countries and China through the KKU Active Recruitment Project 2022. Additionally, partial support was provided by the Advance Smart Computing Laboratory, College of Computing, Khon Kaen University. The authors would like to express their gratitude for the support provided by the Artificial Intelligence Center (AIC), College of Computing, Khon Kaen University, Thailand. The endorsement of the final draft of the manuscript for publication.

REFERENCES

[1] P. E. H. Schwarz *et al.*, "Global Diabetes Survey—An annual report on quality of diabetes care," *Diabetes Research and Clinical Practice*, vol. 100, no. 1, pp. 11–18, Apr. 2013, https://doi.org/10.1016/j.diabres.2012.11.008.

[2] M. I. Campos-Almazán *et al.*, "Computational Methods in Cooperation with Experimental Approaches to Design Protein Tyrosine Phosphatase 1B Inhibitors in Type 2 Diabetes Drug Design: A Review of the Achievements of This Century," *Pharmaceuticals*, vol. 15, no. 7, p. 866, Jul. 2022, https://doi.org/10.3390/ph15070866.

[3] D. Gupta, A. Choudhury, U. Gupta, P. Singh, and M. Prasad, "Computational approach to clinical diagnosis of diabetes disease: a comparative study," *Multimed Tools Appl*, vol. 80, no. 20, pp. 30091–30116, Aug. 2021, https://doi.org/10.1007/s11042-020-10242-8.

[4] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine Learning and Data Mining Methods in Diabetes Research,” *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116, 2017, https://doi.org/10.1016/j.csbj.2016.12.005.

[5] Y. Kumar and S. Gupta, “Deep Transfer Learning Approaches to Predict Glaucoma, Cataract, Choroidal Neovascularization, Diabetic Macular Edema, DRUSEN and Healthy Eyes: An Experimental Review,” *Arch Computat Methods Eng*, vol. 30, no. 1, pp. 521–541, Jan. 2023, https://doi.org/10.1007/s11831-022-09807-7.

[6] A. Nomura, M. Noguchi, M. Kometani, K. Furukawa, and T. Yoneda, “Artificial Intelligence in Current Diabetes Management and Prediction,” *Curr Diab Rep*, vol. 21, no. 12, p. 61, Dec. 2021, https://doi.org/10.1007/s11892-021-01423-2.

[7] M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, “Classification and prediction of diabetes disease using machine learning paradigm,” *Health Inf Sci Syst*, vol. 8, no. 1, p. 7, Dec. 2020, https://doi.org/10.1007/s13755-019-0095-z.

[8] J. Grauslund, “Diabetic retinopathy screening in the emerging era of artificial intelligence,” *Diabetologia*, vol. 65, no. 9, pp. 1415–1423, Sep. 2022, https://doi.org/10.1007/s00125-022-05727-0.

[9] A. Y. Lee *et al.*, “Multicenter, Head-to-Head, Real-World Validation Study of Seven Automated Artificial Intelligence Diabetic Retinopathy Screening Systems,” *Diabetes Care*, vol. 44, no. 5, pp. 1168–1175, May 2021, https://doi.org/10.2337/dc20-1877.

[10] Y. Zhang *et al.*, “Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study,” *BMJ Open Diab Res Care*, vol. 8, no. 1, p. e001596, Oct. 2020, https://doi.org/10.1136/bmjdrc-2020-001596.

[11] J. Shen *et al.*, “An Innovative Artificial Intelligence–Based App for the Diagnosis of Gestational Diabetes Mellitus (GDM-AI): Development Study,” *J Med Internet Res*, vol. 22, no. 9, p. e21573, Sep. 2020, https://doi.org/10.2196/21573.

[12] N. Musacchio *et al.*, “Artificial Intelligence and Big Data in Diabetes Care: A Position Statement of the Italian Association of Medical Diabetologists,” *J Med Internet Res*, vol. 22, no. 6, p. e16922, Jun. 2020, https://doi.org/10.2196/16922.

[13] N. S. Tyler *et al.*, “An artificial intelligence decision support system for the management of type 1 diabetes,” *Nat Metab*, vol. 2, no. 7, pp. 612–619, Jun. 2020, https://doi.org/10.1038/s42255-020-0212-y.

[14] L. M. Polfus *et al.*, “Genetic discovery and risk characterization in type 2 diabetes across diverse populations,” *Human Genetics and Genomics Advances*, vol. 2, no. 2, p. 100029, Apr. 2021, https://doi.org/10.1016/j.xhgg.2021.100029.

[15] L. Muflikhah, F. A. Bachtiar, D. E. Ratnawati, and R. Darmawan, “Improving Performance for Diabetic Nephropathy Detection Using Adaptive Synthetic Sampling Data in Ensemble Method of Machine Learning Algorithms,” *J. Ilm. Tek. Elektro Komput. Dan Inform*, vol. 10, no. 1, p. 123, Feb. 2024, https://doi.org/10.26555/jiteki.v10i1.28107.

[16] H. D. McIntyre, P. Catalano, C. Zhang, G. Desoye, E. R. Mathiesen, and P. Damm, “Gestational diabetes mellitus,” *Nat Rev Dis Primers*, vol. 5, no. 1, p. 47, Jul. 2019, https://doi.org/10.1038/s41572-019-0098-8.

[17] M. D. Bellin and T. B. Dunn, “Transplant strategies for type 1 diabetes: whole pancreas, islet and porcine beta cell therapies,” *Diabetologia*, vol. 63, no. 10, pp. 2049–2056, Oct. 2020, https://doi.org/10.1007/s00125-020-05184-7.

[18] F. Zhang and E. S. Tzanakakis, “Amelioration of Diabetes in a Murine Model upon Transplantation of Pancreatic β-Cells with Optogenetic Control of Cyclic Adenosine Monophosphate,” *ACS Synth. Biol.*, vol. 8, no. 10, pp. 2248–2255, Oct. 2019, https://doi.org/10.1021/acssynbio.9b00262.

[19] J. E. Campbell and C. B. Newgard, “Mechanisms controlling pancreatic islet cell function in insulin secretion,” *Nat Rev Mol Cell Biol*, vol. 22, no. 2, pp. 142–158, Feb. 2021, https://doi.org/10.1038/s41580-020-00317-7.

[20] T. V. Duong *et al.*, “Impact of Percent Body Fat on All-Cause Mortality among Adequate Dialysis Patients with and without Insulin Resistance: A Multi-Center Prospective Cohort Study,” *Nutrients*, vol. 11, no. 6, p. 1304, Jun. 2019, https://doi.org/10.3390/nu11061304.

[21] L. V. Gromova, S. O. Fetissov, and A. A. Gruzdkov, “Mechanisms of Glucose Absorption in the Small Intestine in Health and Metabolic Diseases and Their Role in Appetite Regulation,” *Nutrients*, vol. 13, no. 7, p. 2474, Jul. 2021, https://doi.org/10.3390/nu13072474.

[22] X. Lin *et al.*, “Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025,” *Sci Rep*, vol. 10, no. 1, p. 14790, Sep. 2020, https://doi.org/10.1038/s41598-020-71908-9.

[23] K. W. Lee *et al.*, “Prevalence and risk factors of gestational diabetes mellitus in Asia: a systematic review and meta-analysis,” *BMC Pregnancy Childbirth*, vol. 18, no. 1, p. 494, Dec. 2018, https://doi.org/10.1186/s12884-018-2131-4.

[24] I. IDF, “Indonesia diabetes report 2000 — 2045". Accessed: Dec. 19, 2022. [Online]. Available: https://diabetesatlas.org/data/.

[25] K. Ogurtsova *et al.*, “IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021,” *Diabetes Research and Clinical Practice*, vol. 183, p. 109118, Jan. 2022, https://doi.org/10.1016/j.diabres.2021.109118.

[26] R. L. Thomas, S. Halim, S. Gurudas, S. Sivaprasad, and D. R. Owens, “IDF Diabetes Atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018,” *Diabetes Research and Clinical Practice*, vol. 157, p. 107840, Nov. 2019, https://doi.org/10.1016/j.diabres.2019.107840.

[27] H. Sun *et al.*, "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Research and Clinical Practice*, vol. 183, p. 109119, Jan. 2022, https://doi.org/10.1016/j.diabres.2021.109119.

[28] J. Schofield, J. Ho, and H. Soran, "Cardiovascular Risk in Type 1 Diabetes Mellitus," *Diabetes Ther*, vol. 10, no. 3, pp. 773–789, Jun. 2019, https://doi.org/10.1007/s13300-019-0612-8.

[29] U. Galicia-Garcia *et al.*, "Pathophysiology of Type 2 Diabetes Mellitus," *IJMS*, vol. 21, no. 17, p. 6275, Aug. 2020, https://doi.org/10.3390/ijms21176275.

[30] F. Huang, T. Deng, L. Meng, and X. Ma, "Dietary ginger as a traditional therapy for blood sugar control in patients with type 2 diabetes mellitus: A systematic review and meta-analysis," *Medicine*, vol. 98, no. 13, p. e15054, Mar. 2019, https://doi.org/10.1097/MD.0000000000015054.

[31] A. Helisa, T. H. Saragih, I. Budiman, F. Indriani, and D. Kartini, "Prediction of Post-Operative Survival Expectancy in Thoracic Lung Cancer Surgery Using Extreme Learning Machine and SMOTE," J*ournal of Applied Biomedicine*, vol. 9, no. 2, 2023, https://doi.org/http://dx.doi.org/10.26555/jiteki.v9i2.25973.

[32] A. U. Haq *et al.*, "Intelligent Machine Learning Approach for Effective Recognition of Diabetes in E-Healthcare Using Clinical Data," *Sensors*, vol. 20, no. 9, p. 2649, May 2020, https://doi.org/10.3390/s20092649.

[33] J. Chaki, S. Thillai Ganesh, S. K. Cidham, and S. Ananda Theertan, "Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3204–3225, Jun. 2022, https://doi.org/10.1016/j.jksuci.2020.06.013.

[34] U. Ahmed *et al.*, "Prediction of Diabetes Empowered With Fused Machine Learning," *IEEE Access*, vol. 10, pp. 8529–8538, 2022, https://doi.org/10.1109/ACCESS.2022.3142097.

[35] J. Cadamuro, "Disruption vs. evolution in laboratory medicine. Current challenges and possible strategies, making laboratories and the laboratory specialist profession fit for the future," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 61, no. 4, pp. 558–566, Mar. 2023, https://doi.org/10.1515/cclm-2022-0620.

[36] S. Preston *et al.*, "Toward structuring real-world data: Deep learning for extracting oncology information from clinical text with patient-level supervision," *Patterns*, vol. 4, no. 4, p. 100726, Apr. 2023, https://doi.org/10.1016/j.patter.2023.100726.

[37] V. Baxi, R. Edwards, M. Montalto, and S. Saha, "Digital pathology and artificial intelligence in translational medicine and clinical practice," *Modern Pathology*, vol. 35, no. 1, pp. 23–32, Jan. 2022, https://doi.org/10.1038/s41379-021-00919-2.

[38] D. C. E. Saputra, A. Ma'arif, and K. Sunat, "Optimizing Predictive Performance: Hyperparameter Tuning in Stacked Multi-Kernel Support Vector Machine Random Forest Models for Diabetes Identification," *Journal of Robotics and Control*, vol. 4, no. 6, pp. 896–904, Jan. 2024, https://doi.org/10.18196/jrc.v4i6.20898.

[39] D. C. E. Saputra, Y. Maulana, T. A. Win, R. Phann, and W. Caesarendra, "Implementation of Machine Learning and Deep Learning Models Based on Structural MRI for Identification of Autism Spectrum Disorder," *J. Ilm. Tek. Elektro Komput. dan Inform. JITEKI*, vol. 9, no. 2, 2023, https://doi.org/10.26555/jiteki.v9i2.26094.

[40] L. Cloete, "Diabetes mellitus: an overview of the types, symptoms, complications and management," *Nursing Standard*, vol. 37, no. 1, pp. 61–66, Jan. 2022, https://doi.org/10.7748/ns.2021.e11709.

[41] M. Altman, T. T. K. Huang, and J. Y. Breland, "Design Thinking in Health Care," *Prev. Chronic Dis.*, vol. 15, p. 180128, Sep. 2018, https://doi.org/10.5888/pcd15.180128.

[42] J. Y. Lee and S. W. H. Lee, "Telemedicine Cost–Effectiveness for Diabetes Management: A Systematic Review," *Diabetes Technology & Therapeutics*, vol. 20, no. 7, pp. 492–500, Jul. 2018, https://doi.org/10.1089/dia.2018.0098.

[43] J. Ramesh, R. Aburukba, and A. Sagahyroon, "A remote healthcare monitoring framework for diabetes prediction using machine learning," *Healthc. technol. lett.*, vol. 8, no. 3, pp. 45–57, Jun. 2021, https://doi.org/10.1049/htl2.12010.

[44] M. M. Hassan, M. A. Mamun Billah, M. M. Rahman, S. Zaman, M. M. Hasan Shakil, and J. H. Angon, "Early Predictive Analytics in Healthcare for Diabetes Prediction Using Machine Learning Approach," in *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–5, Jul. 2021, https://doi.org/10.1109/ICCCNT51525.2021.9579799.

[45] S. Sivaranjani, S. Ananya, J. Aravinth, and R. Karthika, "Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction," in *7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2021, pp. 141–146. https://doi.org/10.1109/ICACCS51430.2021.9441935.

[46] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, Dec. 2021, https://doi.org/10.1016/j.icte.2021.02.004.

[47] C. Azad, A. Mehta, D. Mahto, and D. Yadav, "Support Vector Machine based eHealth Cloud System for Diabetes Classification," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, no. 22, p. 164627, May 2020, https://doi.org/10.4108/eai.13-7-2018.164627.

[48] P. Nagaraj and P. Deepalakshmi, "Diabetes Prediction Using Enhanced SVM and Deep Neural Network Learning Techniques: An Algorithmic Approach for Early Screening of Diabetes," *International Journal of Healthcare Information Systems and Informatics*, vol. 16, no. 4, pp. 1–20, Oct. 2021, https://doi.org/10.4018/IJHISI.20211001.oa25.

[49] Z. Xu and Z. Wang, "A Risk Prediction Model for Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XGBoost Ensemble Classifier," in *Eleventh International Conference on Advanced Computational Intelligence (ICACI)*, Jun. 2019, pp. 278–283. https://doi.org/10.1109/ICACI.2019.8778622.

[50] N. A. B. S. Noor, I. Elamvazuthi, and N. Yahya, “Classification of Diabetes Mellitus using Ensemble Algorithms,” in *8th International Conference on Intelligent and Advanced Systems (ICIAS)*, Jul. 2021, pp. 1–6. https://doi.org/10.1109/ICIAS49414.2021.9642508.

[51] P. Palimkar, R. N. Shaw, and A. Ghosh, “Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach,” in *Advanced Computing and Intelligent Technologies*, pp. 219–244, 2021, https://doi.org/10.1007/978-981-16-2164-2_19.

[52] V. K. Kalaiselvi *et al.*, “Analysis of Pima Indian Diabetes Using KNN Classifier and Support Vector Machine Technique,” in *Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, pp. 1376–1380, Aug. 2022, https://doi.org/10.1109/ICICICT54557.2022.9917992.

[53] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006, https://doi.org/10.1016/j.neucom.2005.12.126.

[54] P. Horata, S. Chiewchanwattana, and K. Sunat, “Robust extreme learning machine,” *Neurocomputing*, vol. 102, pp. 31–44, Feb. 2013, https://doi.org/10.1016/j.neucom.2011.12.045.

[55] G. A. Kale and C. Karakuzu, “Multilayer extreme learning machines and their modeling performance on dynamical systems,” *Applied Soft Computing*, vol. 122, p. 108861, Jun. 2022, https://doi.org/10.1016/j.asoc.2022.108861.

[56] U. Gupta and D. Gupta, “Regularized based implicit Lagrangian twin extreme learning machine in primal for pattern classification,” *Int. J. Mach. Learn. & Cyber.*, vol. 12, no. 5, pp. 1311–1342, May 2021, https://doi.org/10.1007/s13042-020-01235-y.

[57] F. Cosmi *et al.*, “Treatment with insulin is associated with worse outcome in patients with chronic heart failure and diabetes: Insulin treatment and prognosis in chronic HF,” *Eur J Heart Fail*, vol. 20, no. 5, pp. 888–895, May 2018, https://doi.org/10.1002/ejhf.1146.

[58] M. J. Davies *et al.*, “Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD),” *Diabetes Care*, vol. 41, no. 12, pp. 2669–2701, Dec. 2018, https://doi.org/10.2337/dci18-0033.

[59] L. C. Khoe, G. Wangge, P. Soewondo, D. L. Tahapary, and I. S. Widyahening, “The implementation of community-based diabetes and hypertension management care program in Indonesia,” *PLoS ONE*, vol. 15, no. 1, p. e0227806, Jan. 2020, https://doi.org/10.1371/journal.pone.0227806.

[60] W. Ling, Y. Huang, Y.-M. Huang, R.-R. Fan, Y. Sui, and H.-L. Zhao, “Global trend of diabetes mortality attributed to vascular complications, 2000–2016,” *Cardiovasc Diabetol*, vol. 19, no. 1, p. 182, Dec. 2020, https://doi.org/10.1186/s12933-020-01159-5.

[61] N. S. Shah *et al.*, “Trends in Gestational Diabetes at First Live Birth by Race and Ethnicity in the US, 2011-2019,” *JAMA*, vol. 326, no. 7, p. 660, Aug. 2021, https://doi.org/10.1001/jama.2021.7217.

[62] L. A. DiMeglio, C. Evans-Molina, and R. A. Oram, “Type 1 diabetes,” *The Lancet*, vol. 391, no. 10138, pp. 2449–2462, Jun. 2018, https://doi.org/10.1016/S0140-6736(18)31320-5.

[63] on behalf of the HAPO Follow-up Study Cooperative Research Group *et al.*, “Maternal glucose levels during pregnancy and childhood adiposity in the Hyperglycemia and Adverse Pregnancy Outcome Follow-up Study,” *Diabetologia*, vol. 62, no. 4, pp. 598–610, Apr. 2019, https://doi.org/10.1007/s00125-018-4809-6.

[64] U. e Laila, K. Mahboob, A. W. Khan, F. Khan, and W. Taekeun, “An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning: An Empirical Study,” *Sensors*, vol. 22, no. 14, p. 5247, Jul. 2022, https://doi.org/10.3390/s22145247.

[65] P. Y. Taser, “Application of Bagging and Boosting Approaches Using Decision Tree-Based Algorithms in Diabetes Risk Prediction,” in *The 7th International Management Information Systems Conference*, p. 6, Mar. 2021, https://doi.org/10.3390/proceedings2021074006.

[66] T. Beghriche, M. Djerioui, Y. Brik, B. Attallah, and S. B. Belhaouari, “An Efficient Prediction System for Diabetes Disease Based on Deep Neural Network,” *Complexity*, vol. 2021, pp. 1–14, Dec. 2021, https://doi.org/10.1155/2021/6053824.

[67] P. Hounguè and A. G. Bigirimana, “Leveraging Pima Dataset to Diabetes Prediction: Case Study of Deep Neural Network,” *JCC*, vol. 10, no. 11, pp. 15–28, 2022, https://doi.org/10.4236/jcc.2022.1011002.

[68] H. Kim, D. H. Lim, and Y. Kim, “Classification and Prediction on the Effects of Nutritional Intake on Overweight/Obesity, Dyslipidemia, Hypertension and Type 2 Diabetes Mellitus Using Deep Learning Model: 4–7th Korea National Health and Nutrition Examination Survey,” *IJERPH*, vol. 18, no. 11, p. 5597, May 2021, https://doi.org/10.3390/ijerph18115597.

[69] R. D. Joshi and C. K. Dhakal, “Predicting Type 2 Diabetes Using Logistic Regression and Machine Learning Approaches,” *IJERPH*, vol. 18, no. 14, p. 7346, Jul. 2021, https://doi.org/10.3390/ijerph18147346.

[70] H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma, “Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction,” *Complex Intell. Syst.*, vol. 8, no. 4, pp. 3073–3087, Aug. 2022, https://doi.org/10.1007/s40747-021-00398-7.

[71] K. C. Howlader *et al.*, “Machine learning models for classification and identification of significant attributes to detect type 2 diabetes,” *Health Inf Sci Syst*, vol. 10, no. 1, p. 2, Feb. 2022, https://doi.org/10.1007/s13755-021-00168-2.

[72] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, and J. Haider, “An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI,” *Sensors*, vol. 22, no. 19, p. 7268, Sep. 2022, https://doi.org/10.3390/s22197268.

BIOGRAPHY OF AUTHORS

Dimas Chaerul Ekty Saputra received his bachelor's degree from the Department of Informatics, Faculty of Industrial Technology, Ahmad Dahlan University, Yogyakarta, Indonesia in 2020, an M.Sc. degree in Department of Biomedical Engineering, Graduate School, Universitas Gadjah Mada, Yogyakarta, Indonesia in 2022, and now he is a PhD candidate in the Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, Thailand. He is a member of the Association for Scientific Computing and Electronics, Engineering (ASCEE) Student Branch Indonesia. His research interests include artificial Intelligence, pattern recognition, machine learning, signal processing, and bioinformatics. Email: dimaschaerulekty.s@kkumail.com, Orcid: 0000-0001-6978-2846.

Elvaro Islami Muryadi received his bachelor's degree in public health, Respati Indonesia University, Jakarta, Indonesia, in 2015. Prior to that, in 2023, he completed a master's program in public health at Prima Indonesia University, Medan, Indonesia, after completing a master's program in business administration at Respati Indonesia University, Jakarta, Indonesia, in 2018. At present, he is engaged in doctoral studies at Khon Kaen University, Thailand, Faculty of Medicine, Department of Community, Occupational, and Family Medicine. Health education, public health, health promotion, and family medicine are among his current research interests. Email: elvaroislamimuryadi.e@kkumail.com, Orcid: 0000-0001-5554-3641.

Raksmey Phann received his M.Sc. degree in Data Science and Artificial Intelligence, Khon Kaen University, Khon Kaen, Thailand, in 2023, and now he is a lecturer in the Department of Applied Mathematics and Statistics, Institute of Technology of Cambodia, Phnom Penh 120408, Cambodia. His research interests include artificial intelligence, natural language processing, machine learning, speech synthesis, text classification, and computer vision. Email: raksmey.p@kkumail.com. Orcid: 0000-0002-5526-8376.

Irianna Futri received her bachelor's degree in information systems from the Nurdin Hamzah School of Computer and Informatics Management in Jambi, Indonesia, in 2011. She worked at The Construction Services Public Works Department Jambi for 1 year as an Administrator for SIPJAKI (Construction Services Development Information System). She also worked as an informatics teacher at the Vocational School of Dharma Bhakti 4 for 3 years and served as an Operator for Dapodik for 7 years. She is pursuing a master's degree in international technology and Innovation Management at the International College, Khon Kaen University, Khon Kaen, Thailand. Her current research interests include advanced care plans and bibliometric analysis. Email: irianna.f@kkumail.com, Orcid: 0009-0003-7100-3583.

Lismawati, graduated from the Prima Midwifery Academy in Jambi, Indonesia, with a degree in advocacy education in 2009, and STIKES Prima Jambi, where she completed her Diploma 4 in midwifery education, from which she graduated in 2011. She subsequently pursued a Master of Public Health with a concentration in reproductive health at Kader Bangsa University, Palembang. Indonesia, from which he graduated in 2015, is presently enrolled at Prima Indonesia University in Medan, where he is pursuing a doctorate in medical sciences.Indonesia and is a lecturer in midwifery at Adiwangsa University of Jambi in Jambi, Indonesia. Email: lismawati101030@gmail.com.