Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 16 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i16.51643

PAPER

Multi-Label Risk Prediction Diabetes Complication Using Machine Learning Models

Nur Rachman Dzakiyullah[1–3](✉), M. A. Burhanuddin[3], Raja Rina Raja Ikram[3], Novanto Yudistira[4], Muhammad Rifqi Fauzi[4], Dwi Joko Purbohadi[5]

[1]Faculty of Computer and Engineering, Department of Information System, Universitas Alma Ata, Yogyakarta, Indonesia

[2]Alma Ata Center for Medical Informatics, Universitas Alma Ata, Yogyakarta, Indonesia

[3]Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka

[4]Faculty of Computer Science, Universitas Brawijaya, Malang, Indonesia

[5]Faculty Engineering, Department of Information Technology, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia

nurrachmandzakiyullah@almaata.ac.id

ABSTRACT

Early diagnosis of diabetic complications based on risk factors is essential but remains understudied, particularly in the context of multi-label classification (MLC). This study leverages data from the behavioral risk factor surveillance system (BRFSS) from 2016 to 2021 to classify seven diabetes complications using MLC techniques combined with multiple machine learning (ML) models. We analyzed 33 variables per dataset year after thorough statistical analysis and preprocessing. Seven ML models were employed: Artificial neural network (ANN), random forest (RF), decision tree (DT), K-nearest neighbors (K-NN), naïve Bayes (NB), support vector machine (SVM), and deep neural network (DNN). We compared two MLC frameworks: problem transformation and algorithm adaptation. The performance of the models was evaluated using several metrics, and feature importance for each complication was analyzed. Our results indicate that the algorithm adaptation framework, particularly with DNN models, outperforms problem transformation. This highlights the potential of this approach for improving classification performance in complex diseases with multiple complications.

KEYWORDS

multi-label classification, risk prediction models, diabetes complication, machine learning (ML), early diagnosis

1 INTRODUCTION

Diabetes is a metabolic disease that can cause complications affecting vital organs of the human body, classified as either microvascular or macrovascular complications, along with associated comorbidities [1]. These complications arise when blood sugar is not adequately controlled, particularly in patients who fail to manage their risk factors through medication or lifestyle changes. According to the international diabetes federation (IDF), 1 in 10 adults aged 20–79 years have diabetes, with projections indicating that the number will rise to 643 million by 2030 and 783 million by 2045 [2]. Additionally, unhealthy lifestyle behaviors, driven by social, economic,

Dzakiyullah, N.R., Burhanuddin, M.A., Raja Ikram, R.R., Yudistira, N., Fauzi, M.R., Purbohadi, D.J. (2024). Multi-Label Risk Prediction Diabetes Complication Using Machine Learning Models. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(16), pp. 66–88. https://doi.org/10.3991/ijoe.v20i16.51643

Article submitted 2024-08-08. Revision uploaded 2024-10-08. Final acceptance 2024-10-08.

or interpersonal pressures, are linked to diabetes risk [3]–[5]. Even in high-income countries such as the United States, the financial burden of diabetes is significant. The American diabetes association (ADA) reported in 2022 that the annual economic cost of diabetes reached $412.9 billion, with $306.6 billion attributed to direct medical expenses and $106.3 billion to indirect costs [6]. Direct medical expenses are expected to increase by 35% from 2012 to 2017, adjusted for inflation [7]. One in four dollars spent on healthcare in the US today goes to helping people with diabetes.

Numerous studies have looked at the connection between lifestyle factors and things such as death, heart disease risk, or controlling risk factors. Some lifestyle choices may increase the risk of microvascular or macrovascular problems, but more research is needed to fully understand this link [3], [8]. Realizing the benefits of starting a healthy lifestyle in the early stages of diabetes could lower the number of people who get serious complications and numerous long-term illnesses, thereby easing the strain on the healthcare system. As a result, early diagnosis of diabetes complications often relies on the doctor's knowledge and experience, which isn't always accurate and can be dangerous. Diabetes is now the leading cause of death, with 76.6% of people using it reporting at least one problem [9]. The need for early diagnosis of diabetes complications is critical, as it can prevent the progression of diseases such as neuropathy, nephropathy, retinopathy, foot disease, and cardiovascular conditions, all of which significantly reduce the quality of life and increase healthcare costs. Traditional diagnostic methods heavily depend on the expertise of healthcare professionals, which can be inaccurate and risky [1]. Therefore, advanced tools that aid in early diagnosis are essential.

Artificial intelligent (AI) today, as a form of digital technology, is one potential solution in healthcare. In particular, machine learning (ML) offers promising solutions for diagnosing and predicting diabetes by leveraging vast datasets [10]–[13]. According to [14], ML as a tool for precision medicine is still in its infancy, with the authors stating, '*we are still learning the strengths and limitations of ML as an approach*.' Previous studies have demonstrated that ML applications in diabetes research can build robust models that effectively identify the relationships between key attributes of diabetes across different dataset settings [15]. For example, ML has been applied in various areas of diabetes research, such as diabetic retinopathy [16]–[20], diabetic foot ulcer detection [21], onset of type 2 diabetes [22]–[24], diabetes complications [25]–[27], and support for self-management [28]. However, comprehensive analysis of such data remains challenging due to limitations in healthcare systems, which cannot provide all types of data, such as tabular, image, video, signal, and text. Despite this, ML offers enormous potential for handling large volumes of data with high-dimensional features and a huge number of examples. ML has also been used extensively in the field of disease prediction [29]–[36]. There are three main ideas in ML for task classification: binary, multiclass, and multi-label classification. These approaches guide how to assign class labels to examples from the problem domain. Several researchers have used these ideas, and they have made a big difference in the field of diabetes prediction [26], [29], [37]–[41]. Furthermore, some studies also focus on comparing the performance of the ML algorithm to achieve the best accuracy [42]–[44], while some researchers are also developing models based on private data or through the electronic health record (EHR) [45], [46]. In fact, most studies usually use the binary classification model, which has good predictive performance for diabetic prediction. A few studies based on public data also have a significant contribution in the field of diabetes prediction by utilizing the Pima Indiana dataset as basis modeling by ML. [22], [47]–[52]. This dataset was collected by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and is stored in the UCI ML repository. The PIMA Indiana dataset is one of the most famous and widely used datasets in many studies. Frequently, binary and multiclass classification is

conducted in their study. However, there is a limited number of studies that apply multi-label classification (MLC), especially in diabetes complication prediction. Several studies in different domains have demonstrated that MLC approaches show great promise in tasks such as text categorization, image classification, automatic annotation for multimedia content, bioinformatics, web mining, rule mining, information retrieval, tag recommendation, and other diverse fields [53]–[58].

Multi-label approaches are a more suitable technique that has not been extensively utilized in diabetes complication prediction studies, as they can more rapidly identify correlations among distinct labels by simultaneously modeling multiple labels [26]. To take advantage of the connections between diabetic complications and enhance their predictive accuracy even further, we thus attempted to employ an MLC model. The prediction of diabetes complications remains challenging due to the multidimensional nature of the disease, which often affects multiple organs simultaneously (microvascular, macrovascular, and comorbidities). Traditionally, ML models used for diabetes complication predictions primarily focus on binary or multi-class classification. However, limited research has explored the MLC approach, which can simultaneously predict multiple complications. This gap in the literature highlights the need for effective modeling strategies to capture interrelated complications. By leveraging large public datasets such as the behavioral risk factor surveillance system (BRFSS), this study aims to explore MLC models and frameworks to enhance predictive accuracy in identifying multiple diabetic complications. MLC allows the simultaneous prediction of multiple related conditions, which may provide a better understanding of the complexity and interactions of diabetic complications than traditional methods. Existing literature indicates that different ML models have varying performance across different datasets. Exploring multiple models in the MLC context is necessary to identify the most robust model. Furthermore, understanding the critical factors contributing to complications is essential for preventive healthcare. This question aims to identify key features using feature importance analysis. In 2019, [59] developed a native ML-based risk prediction model using BRFSS data. The study shows that this public data can be used to design a risk prediction model, and from the experimental results, neural network archive the best performance. Established in 1984, BRFSS is the biggest continually conducted health survey system in the world, with adult interviews conducted in 15 states annually. It collects state-specific data on Americans' health-related risk behaviors, chronic health issues, and preventive service usage [60]. The objective of BRFSS is to identify the risk factors for a variety of human diseases. However, this study provides an exciting opportunity to advance our knowledge of designing multi-label risk prediction diabetes complication models (MLRPDC) by utilizing data from BRFSS and the MLC concept that modeled and predicted each diabetic complication problem separately. Our work has three contributions based on the challenges and evidence discussed:

1. Building risk prediction for diabetes complications by utilizing MLC frameworks and native ML models.
2. We proposed a comparison methodology with performance evaluation for MLC Frameworks.
3. Explaining important factors contributing to diabetes complications.

This paper's organization is as follows: Section 2 discusses material and methods, including elements addressed in this research, experiment setup, and data sets. Section 3 shows the result and discussion with respect to comparing MLC frameworks with several ML, and Section 4 discusses the conclusion of research and presents scope for further study.

2 MATERIAL AND METHODS

This study utilizes seven native models: artificial neural network (ANN), deep neural network (DNN), random forest (RF), decision tree (DT), K-nearest neighbors (KNN), naive Bayes (NB), and support vector machine (SVM). Inspired by the structure and function of the human brain, an ANN is a model of connected neurons arranged in layers, enabling it to learn complex patterns. Multi-layer perceptron (MLP) is commonly used for multi-label classification, with attention mechanisms by improving label dependency capture [61]. A DNN adds extra layers to ANNs, allowing it to learn hierarchical features and perform well on large datasets, making it effective for MLC. A DT model is a flowchart-like model that tests attributes at nodes, with branches representing outcomes and leaves indicating classes. DTs handle both categorical and numerical data and are easy to interpret. Moreover, RF is an ensemble of DTs, combining their predictions for more robust results. KNN classifies instances based on similarity to training data; however, it can be slow with large datasets. NB, a probabilistic method, is efficient for high-dimensional data but struggles with tabular data in multi-label classification. Lastly, SVM finds an optimal hyperplane to separate classes, working well in high-dimensional spaces and adaptable with different kernel functions. To build the MLRPDC models, several steps were undertaken (see Figure 1), which are discussed in the subsequent sections.

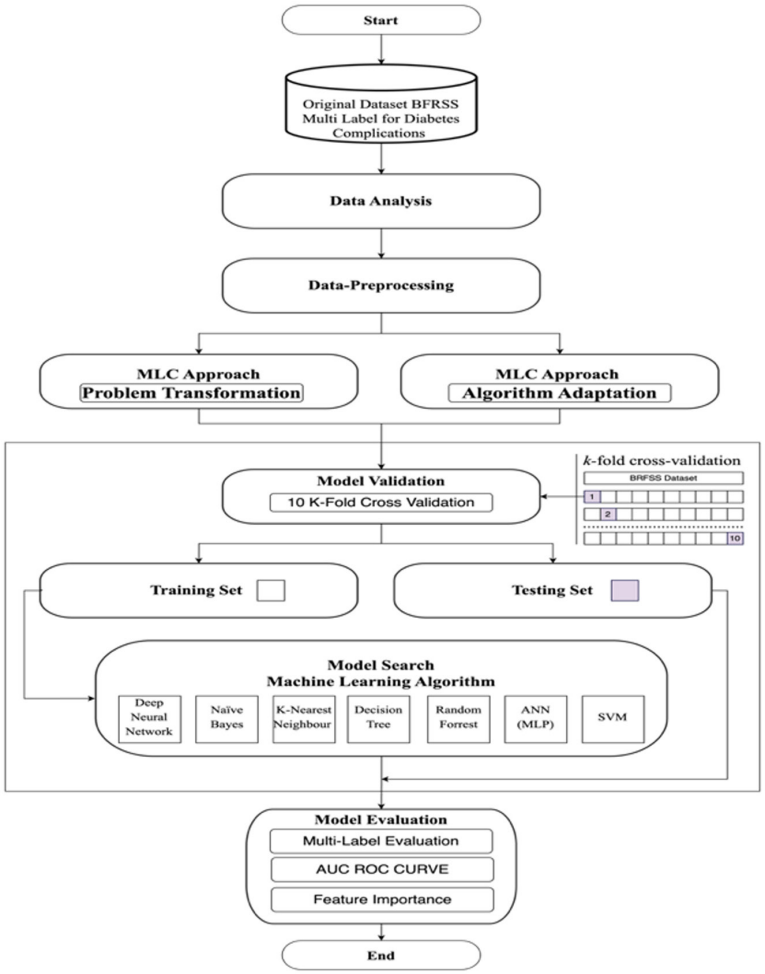

Fig. 1. A framework for multi-label risk prediction diabetes complication

2.1 BRFSS as a diabetes complication dataset

This study utilized BRFSS data from 2016 to 2021, which contains multiple variables for each year (https://www.cdc.gov/brfss/annual_data/annual_data.htm) [60]. In BFRSS 2016–2021 data records, respondents were suffering from diabetes and most of them have associated complications. In the context of diabetes itself, there are three main types of diabetes: type 1, type 2, and gestational. The respondents were diagnosed with pre-diabetes, diabetes, or gestational diabetes based on their answers to the survey questions. However, in this study we only consider 33 variables from each year of BRFSS data. We are extending the study from [59] which considered a number of factors related to type 2 diabetes and this study also considered factors related [62] to diabetes complications in order to build risk prediction complications. Table A1 lists the 33 factors that were investigated as risk diabetes complication factors for this study. The BRFSS dataset contains variables related to diabetic complications and includes respondents with diabetes, prediabetes, and no dependent variable. Other risk factors such as general and mental health status, health care coverage and primary source, and metropolitan status code will also be considered for designing MLRPDC. Independent variables such nephropathy, coronary heart disease, heart attack, stroke, cancer, arthritis, and depression considered as multi-label. To help with personalized type 2 diabetes therapy, we sought to determine which MLC model would be most effective in predicting diabetic complications and can be guiding clinical judgments.

2.2 Data analysis and data pre-processing

Behavioral risk factor surveillance system dataset cannot be useful right away out of the box. In this study, BFFRS dataset was used from 2016 until 2021 with total data 2.632.674 rows x 33 columns and the author fixed the columns by renaming each variable that has been determined X_{input} ($X1 - X26$) and Y_{output} ($Y1 - Y2$). Data cleaning was conducted to detect and remove bad or noisy data, as well as handle missing values. Rows or columns containing NaN values were dropped where data cells remained empty. Then, to handle missing data, this study considered imputing the missing values using mean imputation techniques. In phyton library scikit-learn, specifically, "Iterative Imputer" is a class that implements an imputation algorithm using the MICE (Multiple Imputation by Chained Equations) approach [63]. For each missing feature, this technique builds a regression model. Then, the model's projected values are used to fill in missing feature values until convergence or the maximum number of iterations is reached. The "Iterative Imputer" is initialized with the desired parameters such as missing_values (set to NaN in this case), max_iter (maximum number of iterations), tol (convergence tolerance), n_nearest_features (number of nearest features to use for imputation), and initial_strategy (strategy to use for initializing missing values). After fitting the imputer to Data Frame, the missing values are imputed using the transform method, and the result is stored in the imputed_data DataFrame. The BRFSS dataset also has various issues that must be addressed, such as data values that are irrelevant, if a respondent answered "don't know" or refused to answer. In the BRFSS dataset, age was categorized into groups (1: 31 to 40 y, 2: 41–50 y, 3: 51–60 y, 4: 61–70 y, 5: 71–80 y, 6: >81 y) and mental health data were condensed.

2.3 Multi-label classification frameworks

Machine learning algorithms are trained to predict many labels or categories for a given input instance in MLC. In contrast to single-label classification, each occurrence can be assigned many labels. This study employed varied settings on classification to create MLRPDC using MLC-Problem Transformation (MLC-PT) and MLC-Algorithm Adaptation (MLC-AA). According to [26], [53], [64] MLC involves conversion into different learning issues. First-order, second-order, and high-order techniques are representative. The multi-label learning problem and desired balance between computational efficiency and correlation modelling determine the technique as classified in Table 1.

Table 1. Multi-label learning strategies

Multi-Label Learning Issue	Methods
First-Order	Binary Relevance (BR), Calibrated Label Ranking (CLR), Classifier Chains (CC), Label Powerset (LP), Multi-Label k-Nearest Neighbours (ML-kNN), Multi-Label Random Forest (ML-RF), Multi-Label Artificial Network (ML-ANN), Multi-Label Naïve Bayes (ML-NB), Multi-Label Deep Neural Network (ML-DNN), Multi-Label Decision Tree (ML-DT)
Second-Order	Multi-Label Support Vector Machine (ML-SVM), Calibrated Label Ranking (CLR), Classifier Chains (CC)
Higher-Order	Ensemble methods, Meta-Learning, and other complex techniques that require greater computing complexity

The first-order method is simple and efficient but may overlook label relationships. The second-order strategy captures paired relationships and generalizes well but may miss higher-order correlations. In contrast, the high-order strategy is more comprehensive but requires greater computational complexity and often involves ensemble methods, meta-learning, or other intricate procedures beyond simple adjustments. The choice of strategy depends on the specific multi-label learning problem. Notably, some methods, such as calibrated label ranking (CLR) and CC, can span multiple categories (first-order and second-order) depending on how they manage label dependencies.

Binary relevance (BR) is one of the most straightforward and widely used problem transformation techniques for multi-label classification. The process of transformation on BR for each unique label in the multi-label problem, a binary classifier is trained independently. Each binary classifier is responsible for predicting the presence or absence of a specific label. If there are, for example, three labels (A, B, C) in the original multi-label problem, BR would create three binary classifiers: one for A, one for B, and one for C. The training data for each binary classifier is constructed by considering instances that have the presence or absence of the corresponding label. For instance, for the binary classifier handling label A, instances are labeled as positive if they have label A and negative if they do not. During prediction, each binary classifier produces a binary decision for its corresponding label. The final multi-label prediction is then formed by combining the individual binary decisions. The advantages are BR is conceptually simple and easy to implement and can be combined with various base classifiers, allowing for flexibility in the choice of algorithms. BR assumes that labels are independent, which may not hold true in all cases.

If there is label dependence, other problem transformation methods like CC might be more suitable.

Classifier chains is a problem transformation technique designed to capture label dependencies in multi-label classification. Instead of training a separate binary classifier for each label independently (as in BR), CC creates a chain of classifiers, where each classifier is responsible for predicting one label and considers the predictions of the preceding classifiers in the chain. The order of the chain is often determined based on label dependencies or a predetermined sequence. The first classifier in the chain is trained on the original features and the binary labels for its corresponding label. Each subsequent classifier is trained on the original features along with the binary predictions made by the preceding classifiers in the chain. During prediction, each classifier in the chain produces a binary decision for its corresponding label. The predictions made by preceding classifiers are used as additional features for subsequent classifiers. The final multi-label prediction is formed by combining the binary decisions of all classifiers.

Label Powerset (LP) is a problem transformation technique that treats each unique combination of labels as a distinct class. For a multi-label problem with $\boldsymbol{m}$ labels, LP generates up to $\mathbf{2}^{\boldsymbol{m}}$ possible label combinations. These combinations become separate classes in the transformed problem, where a standard single-label classifier is trained. During prediction, the classifier outputs a label combination, which is then mapped back to the original set of labels for the multi-label prediction. LP effectively captures the joint occurrences of labels, making it suitable when the relationships between label combinations are crucial. However, as the number of labels increases, the number of possible classes grows exponentially, making LP less practical for problems with many labels. In such cases, methods such as BR or CC may be more efficient.

Calibrated label ranking transforms MLC problems by ranking labels based on relevance or likelihood. Instead of predicting labels in a binary fashion or using combinations, CLR ranks each label for an instance according to its relevance or confidence. During training, the model learns to rank relevant labels higher for each instance, and during prediction, it provides scores indicating the relevance of each label. The "calibrated" aspect ensures that the model outputs reflect the relevance of labels more accurately. CLR is particularly effective when a fine-grained understanding of label significance is needed, and it offers more informative predictions when relevance rankings are more meaningful than binary or label combination predictions. Finally, Algorithm adaptation [64], also known as algorithm extension or modification, involves adjusting or extending ML algorithms to address specific challenges or dataset criteria, thereby enhancing performance, robustness, or suitability for a given task. This approach is particularly useful in multi-label classification, where single-label algorithms must be adapted to handle instances with multiple labels. The adaptation process includes modifying algorithms to optimize hyperparameters (such as learning rates, regularization strengths, or kernel parameters) for multi-label datasets, as well as incorporating features that improve the detection of multi-label relationships. This may involve designing elements that account for label dependencies or incorporate domain-specific information. Additionally, algorithm adaptation may involve modeling and leveraging label relationships to improve performance. Ensemble methods, such as bagging or boosting, can further enhance the adapted algorithm's robustness and effectiveness. Adjusting training loss functions to align with MLC goals, such as accounting for label correlations or penalizing specific errors, is another critical aspect. In summary, algorithm adaptation in MLC

enhances performance and applicability by tailoring algorithms or models to better handle multi-label datasets.

2.4 Evaluation metrics for multi-label classification

There is no general agreement about suitable evaluation matric for multi-label classification. The evaluation metrics for multi label learning can be divided into example-based metrics and label-based metrics. In our experimental settings, we conduct several evaluation metrics that has been recommended in the previous study [26], [53], [64] (refer to Table 2):

Table 2. Model evaluation metrics for multi-label classification

Example Based Metric	
Subsets Accuracy	$\frac{1}{p}\sum_{i=1}^{p}\left[\left[h(x_i)=\Upsilon_i\right]\right]$
Hamming Loss	$\frac{1}{p}\sum_{i=1}^{p}\frac{1}{q}\left\|h(x_i)\Delta\Upsilon_i\right\|$
Accuracy (exp)	$\frac{1}{p}\sum_{i=1}^{p}\frac{\left\|\Upsilon_i\cap h(x_i)\right\|}{\left\|\Upsilon_i\cup h(x_i)\right\|}$
Precision (exp)	$\frac{1}{p}\sum_{i=1}^{p}\frac{\left\|\Upsilon_i\cap h(x_i)\right\|}{\left\|h(x_i)\right\|}$
Recall (exp)	$\frac{1}{p}\sum_{i=1}^{p}\frac{\left\|\Upsilon_i\cap h(x_i)\right\|}{\left\|\Upsilon_i\right\|}$
F1-Score (exp)	$\frac{1}{p}\sum_{i=1}^{p}\frac{2\times\left\|\Upsilon_i\cap h(x_i)\right\|}{\left\|\Upsilon_i\right\|+\left\|h(x_i)\right\|}$
One error	$\frac{1}{p}\sum_{i=1}^{p}\left[\left[\left[argmax_{y\in\Upsilon}f(x_i,y)\right]\notin\Upsilon_i\right]\right]$
Coverage	$\frac{1}{p}\sum_{i=1}^{p}max_{y\in\Upsilon_i}rank_f(x_i,y)-1$
Ranking Loss	$\frac{1}{p}\sum_{i=1}^{p}\frac{1}{\left\|\Upsilon_i\right\|\left\|\overline{\Upsilon_i}\right\|}\left\|\left\{(y',y'')\mid f(x_i,y')\le f(x_i,y''),(y',y'')\in\Upsilon_i\times\overline{\Upsilon_i}\right\}\right\|$
Average precision	$\frac{1}{p}\sum_{i=1}^{p}\frac{1}{\left\|\Upsilon_i\right\|}\sum_{y\in\Upsilon_i}\frac{\left\|\left\{y'\mid rank_f(x_i,y')\le rank_f(x_i,y),y'\in\Upsilon_i\right\}\right\|}{rank_f(x_i,y)}$

(Continued)

Table 2. Model evaluation metrics for multi-label classification *(Continued)*

Label-Based Metric	
Accuracy (micro)	$\frac{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{q} TN_j}{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{q} FP_j + \sum_{j=1}^{q} TN_j + \sum_{j=1}^{q} FN_j}$
Accuracy (macro)	$\frac{1}{q}\sum_{j=1}^{q}\frac{TP_j + TN_j}{TP_j + FP_j + TN_j + FN_j}$
Precision (micro)	$\frac{\sum_{j=1}^{q} TP_j}{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{q} FP_j}$
Precision (macro)	$\frac{1}{q}\sum_{i=1}^{q}\frac{TP_j}{TP_j + FP_j}$
Recall (micro)	$\frac{\sum_{j=1}^{q} TP_j}{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{1} FN_j}$
Recall (macro)	$\frac{1}{q}\sum_{i=1}^{q}\frac{TP_j}{TP_j + FN_j}$
F1-Score (micro)	$\frac{2\times\frac{\sum_{j=1}^{q} TP_j}{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{1} FP_j}\times\frac{\sum_{j=1}^{q} TP_j}{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{1} FN_j}}{\frac{\sum_{j=1}^{q} TP_j}{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{1} FP_j} + \frac{\sum_{j=1}^{q} TP_j}{\sum_{j=1}^{q} TP_j + \sum_{j=1}^{1} FN_j}}$
F1-Score (macro)	$\frac{1}{q}\sum_{i=1}^{q}\frac{2\times\frac{TP_j}{TP_j + FP_j}\times\frac{TP_j}{TP_j + FN_j}}{\frac{TP_j}{TP_j + FP_j} + \frac{TP_j}{TP_j + FN_j}}$
AUC (micro)	$\frac{\left\lvert \begin{matrix}\{(x',x'',y',y'') \mid f(x',y') \geq f(x'',y''),(x',y') \\ \in \{(x_i,y) \mid y \in \Upsilon_i, 1 \leq i \leq p\}, (x'',y'') \in \{(x_i,y) \mid y \notin \Upsilon_i, 1 \leq i \leq p\}\}\end{matrix} \right\rvert}{\left\lvert \{(x_i,y) \mid y \in \Upsilon_i, 1 \leq i \leq p\} \right\rvert \left\lvert \{(x_i,y) \mid y \notin \Upsilon_i, 1 \leq i \leq p\} \right\rvert}$
AUC (macro)	$\frac{1}{q}\sum_{i=1}^{q}\frac{\left\lvert \left\{ \begin{matrix}(x'.x'') \mid f(x',y_j) \geq f(x'',y_j),(x',x'') \\ \in \{x_i \mid y_j \in \Upsilon_i, 1 \leq i \leq p\} \times \{x_i \mid y_j \notin \Upsilon_i, 1 \leq i \leq p\}\end{matrix} \right\} \right\rvert}{\left\lvert \{x_i \mid y_j \in \Upsilon_i, 1 \leq i \leq p\} \right\rvert \left\lvert \{x_i \mid y_j \notin \Upsilon_i, 1 \leq i \leq p\} \right\rvert}$

True labels are denoted by Υ_i, while predicted labels for the same sample are denoted by $h(x_i)$. The Δ symbol indicates a symmetric difference between sets, p and q represent examples and class labels, TP_j and FP_j, represent true positives and false positives for label j, and TN_j and FN_j represent true negatives and false negatives for label j. The accuracy and subset accuracy used in MLC are example-based measures computed on the label set, unlike the general accuracy used in binary or multi-class classification. While macro average metrics calculate the metric for each class before averaging them, micro average metrics add up all class contributions to calculate the average metric. According to previous research [40], we evaluate the model using macro and micro criteria.

3 RESULT AND DISCUSSION

We compared the performance of each MLC framework using 20 evaluation metrics (refer to Table 1). Next, we used AUC and ROC curves to evaluate model performance. We labeled the experiments Nephropathy, Coronary Heart Disease, Heart Attack, Stroke, Cancer, Arthritis, and Depression. The dataset was randomly split into the training (80%) and test (20%) sets. All studies then employed 10-fold cross-validation to validate model performance. The objective in this study is to build the risk prediction for diabetes complications by comparing several experimental MLC frameworks with seven ML models and evaluating them using 20 performance evaluation metrics for MLC. The experimental result (refer to Tables 3 and 4) as empirical evidence of the study that become a standpoint on discussion that was divided into two parts: (1) Comparison of ML model on each MLC frameworks; (2) Comparison of the best performance on each MLC frameworks with respect to the ML performance evaluation in the following sub-sections.

3.1 Comparison of ML on MLC framework

To evaluate the performance of different ML methods across each MLC framework, we analyzed their results using 20 evaluation metrics. In PT-BR, three models outperformed others, including native ML methods such as ANN, NB, and DNN. ANN excelled in 13 metrics compared to NB (2 metrics) and DNN (6 metrics). In PT-CC, four models outperformed native methods like ANN, RF, NB, and DNN, with DNN leading in 10 metrics, followed by RF (3 metrics), NB (3 metrics), and ANN (10 metrics). Similarly, in PT-LP, four models outperformed RF, SVM, NB, and DNN, with RF excelling in 13 metrics, compared to SVM (1 metric), NB (1 metric), and DNN (5 metrics). In PT-CLR, four models outperformed ANN, RF, NB, and DNN, with ANN leading in 10 metrics, RF in 2 metrics, NB in 5 metrics, and DNN in 5 metrics. Finally, in the AA framework, four models outperformed ANN, RF, NB, and DNN, with DNN excelling in 13 metrics, followed by RF (1 metric), NB (2 metrics), and ANN (6 metrics). Overall, the results indicate that ANN, DNN, RF, NB, and SVM demonstrate significant performance across MLC frameworks.

3.2 Comparison the best performance on MLC frameworks

As previously discussed, ANN, DNN, RF, NB, and SVM show significant performance across MLC frameworks, with DNN and NB being the most effective in

handling all MLC problems. The ANN model performed exceptionally well in PT-BR, PT-CC, PT-CLR, and AA, while the RF model excelled in PT-CC, PT-LP, PT-CLR and AA. SVM, however, showed notable performance only in PT-CLR. When evaluating the best performance metrics, ANN stood out in PT-BR and PT-CLR, DNN in PT-CC and AA, and RF in PT-LP.

To determine the overall best performance across all MLC frameworks, we conducted a ranking procedure based on each model's performance. The steps for this ranking procedure are as follows:

1. After the experiment, results are obtained based on performance for MLC frameworks. (refer to Tables 3 and 4)
2. Finding the best results or best model on each MLC metric.
3. Selecting the best metric on ML model from 20 predetermined metrics, the best metric is accumulated as a reference in ranking. Then, determined the five best models in each Framework. (See section 3.2)
4. Calculating the average of the accumulated MLC metrics for each model from 5 Frameworks. To calculated average metrics, follow Eq. 1:

$$Average\ Metric = \frac{\sum_{i=1}^{n} Metric_i}{n} \tag{1}$$

   Where:
   - n is the number of models.
   - $Metric_i$ indicates total metric values for the ith model.
5. Perform ranking and get the best ML model for all frameworks.

The findings from the ranking procedure, based on the average metric values, indicate that DNN (7.6) is the best model for all MLC frameworks, followed by ANN (7.0), RF (3.8), NB (2.6), and lastly, SVM (0.2). These results were obtained by calculating the average of accumulated MLC metrics for each model across five frameworks. Equation 1 was applied to each model, yielding an average metric value. As previously mentioned, the ranking results show that DNN is the top-performing model, with ANN in second place for MLC frameworks. This provides empirical evidence that neural network-based models are more robust compared to other ML models in MLC frameworks. However, it is worth noting that the PT-LP-DNN framework exhibited the worst performance, with 15 of the 20-performance metrics ranking lowest. This poor performance may be attributed to the large number of labels and limited samples, which likely exacerbated the issue of class imbalance in the dataset, as noted by [26].

3.3 Result of area under the receiver operating characteristic curve

To illustrate the performance of the best method, the AA-DNN model for predicting diabetes complications, we utilized AUC and ROC curves. These curves demonstrate the trade-off between recall (sensitivity or true positive rate) and the false-positive rate, providing a comprehensive view of model performance. Additionally, we used the precision-recall curve, which highlights the precision of the model relative to recall, particularly in cases of imbalanced datasets. Notably, the AA-DNN model achieved the highest AUROC, indicating its superior ability to distinguish between classes (see Figure 2). These results validate the AA-DNN model as a highly effective tool for risk prediction in diabetic complications.

Table 3. Result example based metric

MLC Framework	Classifier	Example Based Matric									
		Sub-Set Accuracy ↑	Hamming Loss ↓	Accuracy (Exp) ↑	Precision (Exp) ↑	Recall (Exp) ↑	F1-Score (Exp) ↑	One Error ↓	Coverage ↓	Ranking Loss ↓	Average Precision ↑
PT-BR	**ANN**	**0.4164**	**0.1269**	**0.8620**	**0.8909**	**0.9621**	**0.9114**	**0.0727**	6.5988	0.3212	0.8707
	RF	0.3730	0.1383	0.8499	0.8913	0.9472	0.9008	0.0728	6.6106	0.3194	0.8664
	DT	0.2270	0.2023	0.7778	0.8836	0.8621	0.8538	0.0757	6.7645	0.3512	0.8561
	K-NN	0.3571	0.1450	0.8426	0.8885	0.9410	0.8961	0.0730	6.6461	0.3318	0.8635
	NB	0.3346	0.2294	0.7101	0.7790	0.7780	0.8448	0.0772	6.8327	0.4166	**0.8820**
	SVM	0.0012	0.6551	0.2353	0.6332	0.2464	0.3531	0.0946	6.9918	0.4666	0.8406
	DNN	0.4149	0.1272	0.8615	0.8926	0.9595	0.9103	0.0727	**6.5889**	**0.3129**	0.8713
PT-CC	**ANN**	0.4163	0.1274	**0.8617**	0.8893	**0.9637**	**0.9116**	**0.0727**	6.6142	0.3303	0.8698
	RF	0.3729	0.1382	0.8500	**0.8909**	0.9479	0.9011	0.0728	6.6176	**0.3229**	0.8662
	DT	0.2322	0.2037	0.7760	0.8833	0.8599	0.8526	0.0757	6.7571	0.3502	0.8560
	K-NN	0.3599	0.1442	0.8439	0.8870	0.9446	0.8974	0.0730	6.6531	0.3384	0.8627
	NB	0.3439	0.2316	0.7072	0.7265	0.7926	0.8481	**0.0727**	6.8813	0.4784	**0.8747**
	SVM	0.0088	0.6984	0.1693	0.5055	0.1762	0.2555	0.0768	6.9916	0.4866	0.8402
	DNN	**0.4165**	**0.1273**	**0.8617**	0.8897	0.9631	0.9115	**0.0727**	**6.6131**	0.3283	0.8700
PT-LP	ANN	0.1403	0.2091	0.7734	0.9241	0.8313	0.7118	0.0727	6.5755	0.1666	0.8365
	RF	**0.4141**	**0.1288**	**0.8609**	0.8846	**0.9680**	**0.9122**	0.0727	6.6557	0.3558	**0.8671**
	DT	0.2492	0.1895	0.7920	0.8851	0.8798	0.8619	0.0751	6.7392	0.3460	0.8565
	K-NN	0.3398	0.1501	0.8375	0.8899	0.9340	0.8891	0.0729	6.6520	0.3242	0.8604
	NB	0.1210	0.5356	0.4194	0.7620	0.4559	0.5730	0.0951	6.9692	0.5092	0.8505
	SVM	0.1672	0.2130	0.7694	0.9065	0.8409	0.7984	**0.0719**	6.6645	0.2447	0.8435
	DNN	0.1175	0.2261	0.7542	**0.9274**	0.8046	0.6867	0.0727	**6.6124**	**0.1569**	0.8365
PT-CLR	**ANN**	**0.4166**	0.1271	**0.8621**	0.8889	**0.9645**	**0.9120**	**0.0727**	**6.6108**	0.3306	0.8699
	RF	0.3749	0.1380	0.8500	0.8916	0.9472	0.9009	**0.0727**	6.6127	**0.3188**	0.8666
	DT	0.2685	0.1829	0.7953	0.8931	0.8783	0.8632	0.0738	6.7277	0.3257	0.8606
	KNN	0.3573	0.1450	0.8419	0.8908	0.9381	0.8952	0.0730	6.6452	0.3249	0.8644
	NB	0.3328	0.2317	0.7074	0.7763	0.7751	0.8435	0.0772	6.8372	0.4189	**0.8821**
	SVM	0.2308	0.3029	0.6703	0.7518	0.7911	0.8016	0.0845	6.9596	0.5154	0.8351
	DNN	0.4160	**0.1269**	0.8620	**0.8910**	0.9620	0.9113	**0.0727**	6.5965	0.3203	0.8709
AA	**ANN**	0.4150	0.1273	0.8365	**0.9999**	**0.9996**	0.9109	**0.0584**	6.2347	**0.0823**	0.9343
	RF	0.3733	0.1381	**0.8500**	0.8914	0.9472	0.9008	0.0728	6.6099	0.3190	0.8665
	DT	0.2271	0.2021	0.7780	0.8836	0.8622	0.8539	0.0756	6.7643	0.3511	0.8562
	KNN	0.3567	0.1450	0.8426	0.8885	0.9411	0.8962	0.0730	6.6466	0.3320	0.8635
	NB	0.3346	0.2294	0.7101	0.7790	0.7780	0.8448	0.0773	6.8328	0.4166	0.8820
	SVM	0.0227	0.5799	0.3289	0.6652	0.3533	0.4820	0.0941	6.9687	0.4461	0.8424
	DNN	**0.4156**	**0.1272**	0.8365	0.9989	**0.9996**	**0.9113**	0.0594	**6.2342**	0.0827	**0.9345**

Notes: ↑ and ↓ show the best performance.

Table 4. Result label based matric

MLC Framework	Classifier	Label Based Matric									
		Accuracy (Micro) ↑	Accuracy (Macro) ↑	Precision (Micro) ↑	Precision (Macro) ↑	Recall (Micro) ↑	Recall (Macro) ↑	F1-Score (Micro) ↑	F1-Score (Macro) ↑	AUC (Micro) ↑	AUC (Macro) ↑
PT-BR	**ANN**	**0.8731**	**0.8731**	0.8935	0.8785	**0.9630**	**0.9477**	**0.9270**	**0.9114**	0.6879	0.5569
	RF	0.8617	0.8617	0.8936	0.8755	0.9475	0.9278	0.9198	0.9008	0.6851	0.5547
	DT	0.7977	0.7977	0.8885	0.8645	0.8670	0.8434	0.8776	0.8538	0.6550	0.5648
	K-NN	0.8550	0.8550	0.8908	0.8718	0.9421	0.9221	0.9158	0.8961	0.6756	0.5556
	NB	0.7706	0.7706	0.9072	0.8952	0.8084	0.8026	0.8550	0.8448	0.6926	**0.6761**
	SVM	0.3449	0.3449	0.8873	0.8419	0.2482	0.2388	0.3872	0.3531	0.5441	0.5139
	DNN	0.8728	0.8728	**0.8953**	**0.8801**	0.9602	0.9438	0.9266	0.9103	**0.6928**	0.5573
PT-CC	**ANN**	0.8726	0.8726	0.8918	0.8770	**0.9648**	**0.9502**	**0.9268**	**0.9116**	0.6828	0.5556
	RF	0.8618	0.8618	**0.8931**	0.8750	0.9483	0.9289	0.9199	0.9011	0.6837	0.5534
	DT	0.7963	0.7963	0.8887	0.8645	0.8648	0.8410	0.8766	0.8526	0.6553	0.5647
	K-NN	0.8558	0.8558	0.8890	0.8703	0.9457	0.9265	0.9165	0.8974	0.6708	0.5529
	NB	0.7684	0.7684	0.8907	0.8830	0.8243	0.8235	0.8562	0.8481	0.6533	**0.6624**
	SVM	0.3016	0.3016	0.9304	0.8546	0.1790	0.1657	0.2943	0.2555	0.5539	0.5228
	DNN	**0.8727**	**0.8727**	0.8923	**0.8773**	0.9642	0.9494	**0.9268**	0.9115	**0.6842**	0.5560
PT-LP	ANN	0.7909	0.7909	0.9240	0.6852	0.8191	0.7429	0.8672	0.7118	0.7333	0.5000
	RF	**0.8712**	**0.8712**	0.8870	**0.8725**	**0.9696**	**0.9571**	**0.9265**	**0.9122**	0.6686	0.5522
	DT	0.8105	0.8105	0.8891	0.8649	0.8836	0.8590	0.8864	0.8619	0.6598	0.5610
	K-NN	0.8499	0.8499	0.8915	0.8696	0.9343	0.9102	0.9124	0.8891	0.6760	0.5476
	NB	0.4644	0.4644	0.8016	0.8691	0.4777	0.5172	0.5977	0.5730	0.4367	**0.5636**
	SVM	0.7870	0.7870	0.9047	0.8508	0.8335	0.7804	0.8674	0.7984	0.6914	0.5269
	DNN	0.7739	0.7739	**0.9273**	0.6624	0.7918	0.7143	0.8542	0.6867	**0.7371**	0.5000
PT-CLR	**ANN**	0.8729	0.8729	0.8916	0.8769	**0.9655**	**0.9514**	**0.9271**	**0.9120**	0.6822	0.5562
	RF	0.8620	0.8620	0.8939	0.8758	0.9475	0.9278	0.9199	0.9009	0.6858	0.5549
	DT	0.8171	0.8171	0.8973	0.8723	0.8823	0.8551	0.8897	0.8632	0.6828	0.5709
	KNN	0.8550	0.8550	0.8929	0.8735	0.9393	0.9182	0.9155	0.8952	0.6815	0.5574
	NB	0.7683	0.7683	**0.9069**	**0.8953**	0.8058	0.8005	0.8534	0.8435	**0.6912**	**0.6772**
	SVM	0.6971	0.6971	0.8398	0.8341	0.7882	0.7829	0.8129	0.8016	0.5095	0.4882
	DNN	**0.8731**	**0.8731**	0.8937	0.8788	0.9628	0.9476	0.9270	0.9113	0.6882	0.5569
AA	**ANN**	0.8727	0.8727	0.8937	0.8785	0.9622	0.9466	0.9267	0.9109	**0.8738**	**0.7935**
	RF	0.8619	0.8619	0.8938	0.8757	0.9474	0.9277	0.9198	0.9008	0.6857	0.5550
	DT	0.7979	0.7979	0.8886	0.8646	0.8671	0.8435	0.8777	0.8539	0.6553	0.5654
	KNN	0.8550	0.8550	0.8908	0.8718	0.9422	0.9222	0.9158	0.8962	0.6756	0.5557
	NB	0.7706	0.7706	**0.9072**	**0.8951**	0.8085	0.8026	0.8550	0.8448	0.6925	0.6761
	SVM	0.4201	0.4201	0.8837	0.8537	0.3534	0.3427	0.5038	0.4820	0.5576	0.5121
	DNN	**0.8728**	**0.8728**	0.8930	0.8780	**0.9634**	**0.9483**	**0.9269**	**0.9113**	0.8737	**0.7935**

Notes: ↑ and ↓ showing the best performance.

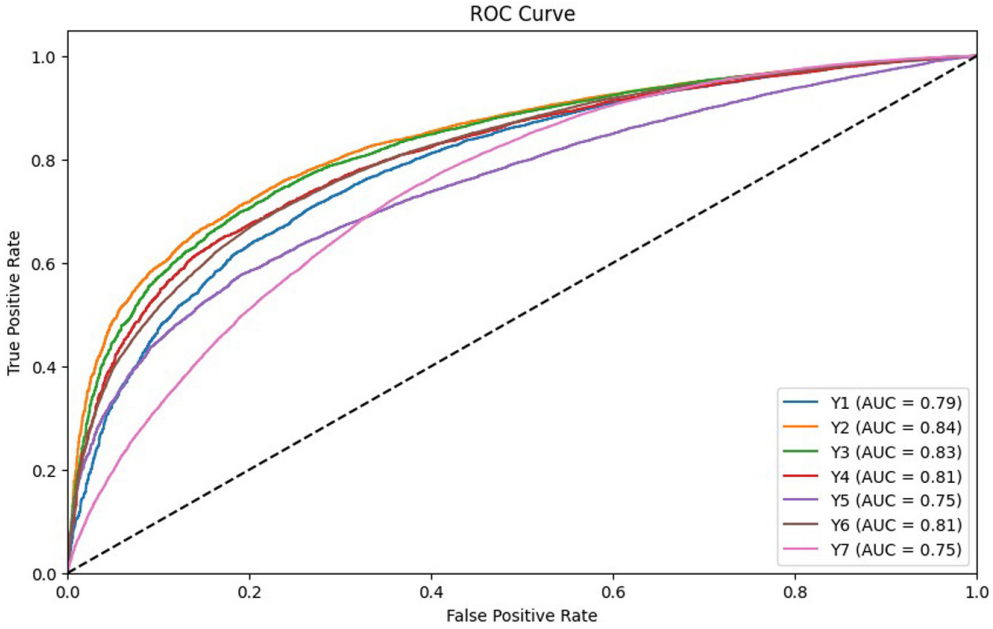

Fig. 2. AUROC of AA-DNN model

Coronary heart disease (Y2) exhibited the highest AUROC among all diabetes complications, followed by cancer (Y5) and depression (Y7). These results reinforce the strong performance of the AA-DNN model, which achieved an AUC macro value of 0.7935. This demonstrates the model's overall effectiveness in predicting multiple diabetes-related complications.

3.4 Result of feature importance

To identify the key factors that significantly contribute to the development of diabetes complications, we utilized the AA-DNN model in this study. The model highlighted several factors with strong correlations to complications, as shown in Table 5. These factors were identified as having a significant impact on the progression of diabetes-related conditions, providing valuable insights into the underlying contributors to these complications.

Table 5. Top10 key features in prediction

Nephropathy (Y1)	Coronary Heart Disease (Y2)	Heart Attack (Y3)	Stroke (Y4)	Cancer (Y5)	Arthritis (Y6)	Depressive Disorder (Y7)
GENHLTH (X1)	_AGEG5YR (X21)	GENHLTH (X1)	GENHLTH (X1)	_AGEG5YR (X21)	_AGEG5YR (X21)	DECIDE (X16)
_AGEG5YR (X21)	GENHLTH (X1)	_AGEG5YR (X21)	_AGEG5YR (X21)	GENHLTH (X1)	GENHLTH (X1)	MENTHLTH (X2)
EMPLOY1 (X11)	_SEX (X17)	_SEX (X17)	EMPLOY1 (X11)	_SEX (X17)	_SEX (X17)	_SEX (X17)
DIABETE4 (X23)	DIABETE4 (X23)	DIABETE4 (X23)	DECIDE (X16)	EMPLOY1 (X11)	EMPLOY1 (X11)	EMPLOY1 (X11)
CHECKUP1 (X4)	EMPLOY1 (X11)	_SMOKER3 (X6)	DIABETE4 (X23)	MARITAL (X9)	_BMI5CAT (X20)	GENHLTH (X1)
BLIND (X15)	_SMOKER3 (X6)	EMPLOY1 (X11)	INCOME2 (X12)	_BMI5CAT (X20)	DECIDE (X16)	_SMOKER3 (X6)
DECIDE (X16)	INCOME2 (X12)	INCOME2 (X12)	BLIND (X15)	_SMOKER3 (X6)	CHECKUP1 (X4)	_AGEG5YR (X21)
_EDUCAG (X13)	BLIND (X15)	CHECKUP1 (X4)	_SMOKER3 (X6)	_EDUCAG (X13)	_SMOKER3 (X6)	_EDUCAG (X13)
DEAF (X22)	CHECKUP1 (X4)	BLIND (X15)	CHECKUP1 (X4)	CHECKUP1 (X4)	_RACE (X19)	_RACE (X19)
MSCODE (X14)	PREDIAB1 (X25)	MARITAL (X9)	MARITAL (X9)	RENTHOM1 (X10)	MARITAL (X9)	_BMI5CAT (X20)

This study presents 10 key features (refer to Table 5) that have been used to validate the AA-DNN model, achieving better performance as demonstrated in the previous analysis of MLC performance, where higher rankings correspond to higher SHAP (Shapley Additive Explanations) values [65]. Many of these features align with established predictive factors; for example, X1, X11, and X21, which are significant in the context of diabetes complications, were consistent with known behavioral risk factors in medical research. To the best of our knowledge, the MLC model can effectively handle the multifactorial correlations of diabetic complications synchronously with high predictive performance. Among the evaluated MLC frameworks, AA-DNN outperformed other PT-MLC when assessed using various performance metrics. As mentioned earlier, the AA-DNN model demonstrated superior performance based on overall MLC metrics, achieving the highest AUROC value of 0.84. Notably, the AA-DNN model showed significant improvement in predicting coronary heart disease (Y2), a crucial predictor of diabetes complications, while cancer (Y5) and depressive disorder (Y7) exhibited the lowest AUROC value of 0.75, reinforcing the accuracy of our model.

4 CONCLUSION AND FUTURE WORK

This study employed MLC frameworks and ML models to predict diabetic complications using public health data. Our findings suggest that the AA framework, particularly with the DNN, is the most effective for predicting multiple diabetic complications. This work demonstrates the potential of ML models in healthcare, specifically in the early diagnosis of complex conditions like diabetes. However, there are certain limitations, as this model is not based on clinical patient data, despite diabetes being a metabolic disease where many studies utilize clinical data or biomarker datasets. A deeper understanding of health studies and the medical context of data is necessary for designing models to diagnose diabetic complications accurately. Additionally, addressing the challenge of imbalanced datasets is crucial, as it enhances classification performance. Future work will explore the feasibility of DL methods that incorporate higher-order strategies for MLC frameworks, with a particular focus on the AA framework.

5 ACKNOWLEDGEMENT

The authors express their gratitude to Universitas Alma Ata for funding this study. Appreciation is also extended to the BIOCORE Research Group, the Center for Advanced Computing Technology (C-ACT), Fakulti Teknologi Maklumat dan Komunikasi (FTMK), the Centre for Research and Innovation Management (CRIM) at Universiti Teknikal Malaysia Melaka (UTeM), and Universitas Brawijaya for providing the necessary facilities and support. All authors have reviewed and approved the final version of the manuscript.

6 CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Nur Rachman Dzakiyullah: Conceptualization, Investigation, Data curation, Methodology, Formal analysis, and Writing original draft. M. A. Burhanuddin: Conceptualization, Data curation, Formal analysis, Supervision, Writing–review

and editing. Raja Rina Raja Ikram: Conceptualization, Data curation, Formal analysis, Supervision, Writing–review and editing. Novanto Yudistira: Conceptualization, Data curation, Formal analysis, Supervision, Writing–review and editing. Muhammad Rifqi Fauzi: Data curation, Programming, Visualization, and Writing–original draft. Dwi Joko Purbohadi: Conceptualization, Data curation, Formal analysis, and Supervision.

7 DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8 REFERENCES

[1] E. Bonora and R. A. DeFronzo (Eds.), *Diabetes Complications, Comorbidities and Related Disorders*. Cham: Springer, 2020. https://doi.org/10.1007/978-3-319-27316-7

[2] International Diabetes Federation, *IDF Diabetes Atlas 2021 (10th Ed.)*, 2021.

[3] Y. V. Chudasama and K. Khunti, "Healthy lifestyle choices and microvascular complications: New insights into diabetes management," *PLoS Med.*, vol. 20, no. 1, p. e1004152, 2023. https://doi.org/10.1371/journal.pmed.1004152

[4] M. Ischaq Nabil Asshiddiqi, K. Yodchai, and P. Taniwattananon, "Predictors of diabetes distress among older persons with type 2 diabetes mellitus in Indonesia," *Journal of Research in Nursing*, vol. 26, no. 4, pp. 307–317, 2021. https://doi.org/10.1177/1744987120943936

[5] E. Afifah, D. S. Nurdiati, H. Hadi, Z. M. Sofro, and A. H. Sadewa, "Social nervous exercise intervention and its association with fasting blood glucose on diabetes mellitus gestational," *Open Access Maced. J. Med. Sci.*, vol. 10, no. B, pp. 129–136, 2022. https://doi.org/10.3889/oamjms.2022.7893

[6] Virginia Cramer, "New American diabetes association report finds annual costs of diabetes to be $412.9 Billion," 2023. [Online]. Available: https://diabetes.org/newsroom/press-releases/new-american-diabetes-association-report-finds-annual-costs-diabetes-be [Accessed: Nov. 27, 2023].

[7] E. D. Parker *et al.*, "Economic costs of diabetes in the U.S. in 2022," *Diabetes Care*, vol. 47, no. 1, pp. 26–43, 2023. https://doi.org/10.2337/dci23-0085

[8] A. Mansour, M. Mousa, D. Abdelmannan, G. Tay, A. Hassoun, and H. Alsafar, "Microvascular and macrovascular complications of type 2 diabetes mellitus: Exome wide association analyses," *Front Endocrinol (Lausanne)*, vol. 14, p. 1143067, 2023. https://doi.org/10.3389/fendo.2023.1143067

[9] H. Hu *et al.*, "A systematic review of the direct economic burden of type 2 diabetes in China," *Diabetes Therapy*, vol. 6, pp. 7–16, 2015. https://doi.org/10.1007/s13300-015-0096-0

[10] J. M. Puaschunder, "The potential for artificial intelligence in healthcare," *SSRN*, 2020. https://doi.org/10.2139/ssrn.3525037

[11] F. Jiang *et al.*, "Artificial intelligence in healthcare: Past, present and future," *Stroke Vasc. Neurol.*, vol. 2, no. 4, 2017. https://doi.org/10.1136/svn-2017-000101

[12] V. Jaiswal, A. Negi, and T. Pal, "A review on current advances in machine learning based diabetes prediction," *Prim. Care Diabetes*, vol. 15, no. 3, pp. 435–443, 2021. https://doi.org/10.1016/j.pcd.2021.02.005

[13] N. Jayanthi, B. V. Babu, and N. S. Rao, "Survey on clinical prediction models for diabetes prediction," *J. Big Data*, vol. 4, 2017. https://doi.org/10.1186/s40537-017-0082-7

[14] D. Plant and A. Barton, "Machine learning in precision medicine: Lessons to learn," *Nat. Rev. Rheumatol.*, vol. 17, pp. 5–6, 2021. https://doi.org/10.1038/s41584-020-00538-2

[15] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017. https://doi.org/10.1016/j.csbj.2016.12.005

[16] S. K. Somasundaram and P. Alli, "A machine learning ensemble classifier for early prediction of diabetic retinopathy," *J. Med. Syst.*, vol. 41, 2017. https://doi.org/10.1007/s10916-017-0853-x

[17] A. Garifullin, L. Lensu, and H. Uusitalo, "Deep Bayesian baseline for segmenting diabetic retinopathy lesions: Advances and challenges," *Comput. Biol. Med.*, vol. 136, p. 104725, 2021. https://doi.org/10.1016/j.compbiomed.2021.104725

[18] M. Aatila, M. Lachgar, H. Hrimech, and A. Kartit, "Diabetic retinopathy classification using ResNet50 and VGG-16 pretrained networks," *International Journal of Computer Engineering and Data Science (IJCEDS)*, vol. 1, no. 1 SE-Articles, pp. 1–7, 2021.

[19] G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. Teatini Salles, "Diabetic retinopathy detection using red lesion localization and convolutional neural networks," *Comput. Biol. Med.*, vol. 116, p. 103537, 2020. https://doi.org/10.1016/j.compbiomed.2019.103537

[20] G. Alfian *et al.*, "Deep neural network for predicting diabetic retinopathy from risk factors," *Mathematics*, vol. 8, no. 9, p. 1620, 2020. https://doi.org/10.3390/math8091620

[21] M. H. Yap *et al.*, "Deep learning in diabetic foot ulcers detection: A comprehensive evaluation," *Comput. Biol. Med.*, vol. 135, p. 104596, 2021. https://doi.org/10.1016/j.compbiomed.2021.104596

[22] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," *J. Diabetes. Metab. Disord.*, vol. 19, pp. 391–403, 2020. https://doi.org/10.1007/s40200-020-00520-5

[23] J.-H. Wu *et al.*, "Risk prediction of type 2 diabetes in steel workers based on convolutional neural network," *Neural Comput. Appl.*, vol. 32, pp. 9683–9698, 2020. https://doi.org/10.1007/s00521-019-04489-y

[24] M. Rahman, D. Islam, R. J. Mukti, and I. Saha, "A deep learning approach based on convolutional LSTM for detecting diabetes," *Comput. Biol. Chem.*, vol. 88, p. 107329, 2020. https://doi.org/10.1016/j.compbiolchem.2020.107329

[25] S. Farzi, S. Kianian, and I. Rastkhadive, "Predicting serious diabetic complications using hidden pattern detection," in *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, 2017, pp. 0063–0068. https://doi.org/10.1109/KBEI.2017.8324885

[26] L. Zhou, X. Zheng, D. Yang, Y. Wang, X. Bai, and X. Ye, "Application of multi-label classification models for the diagnosis of diabetic complications," *BMC Med. Inform. Decis. Mak.*, vol. 21, 2021. https://doi.org/10.1186/s12911-021-01525-7

[27] B. Ljubic *et al.*, "Predicting complications of diabetes mellitus using advanced machine learning algorithms," *Journal of the American Medical Informatics Association*, vol. 27, no. 9, pp. 1343–1351, 2020. https://doi.org/10.1093/jamia/ocaa120

[28] 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, 2022. https://doi.org/10.1016/j.jksuci.2020.06.013

[29] O. Rabie, D. Alghazzawi, J. Asghar, F. K. Saddozai, and M. Z. Asghar, "A decision support system for diagnosing diabetes using deep neural network," *Front Public Health*, vol. 10, 2022. https://doi.org/10.3389/fpubh.2022.861062

[30] T. P. Anish and P. M. J. Prathap, "Biomedical signal processing and control Hsc-attention multi-disease prediction Net: A heuristic-assisted hybrid serial cascaded Attention-based network with ensemble feature selection process for multi-disease prediction," *Biomed Signal Process Control*, vol. 89, p. 105676, 2024. https://doi.org/10.1016/j.bspc.2023.105676

[31] F. Yang, H. Wang, S. Wei, G. Sun, Y. Chen, and L. Tao, "Multi-model adaptive fusion-based graph network for Alzheimer's disease prediction," *Comput. Biol. Med.*, vol. 153, p. 106518, 2023. https://doi.org/10.1016/j.compbiomed.2022.106518

[32] K. Li *et al.*, "DeAF: A multimodal deep learning framework for disease prediction," *Comput. Biol. Med.*, vol. 156, p. 106715, 2023. https://doi.org/10.1016/j.compbiomed.2023.106715

[33] S. A. Siddiqui, A. Ahmad, and N. Fatima, "IoT-based disease prediction using machine learning," *Computers and Electrical Engineering*, vol. 108, p. 108675, 2023. https://doi.org/10.1016/j.compeleceng.2023.108675

[34] A. Ayodele, A. Adetunla, and E. Akinlabi, "Prediction of depression severity and personalised risk factors using machine learning on multimodal data," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 9, pp. 130–143, 2024. https://doi.org/10.3991/ijoe.v20i09.47581

[35] A. Gago, J. M. Aguirre, and L. Wong, "Machine learning system for the effective diagnosis and survival prediction of breast cancer patients," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 2, pp. 95–113, 2024. https://doi.org/10.3991/ijoe.v20i02.42883

[36] S. Hamida, D. Lamrani, M. A. Bouqentar, O. El Gannour, and B. Cherradi, "An integrated multimodal deep learning framework for accurate skin disease classification," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 2, pp. 78–94, 2024. https://doi.org/10.3991/ijoe.v20i02.43795

[37] 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," *Int. J. Environ. Res. Public Health.*, vol. 18, no. 11, p. 5597, 2021. https://doi.org/10.3390/ijerph18115597

[38] H. Zhou, R. Myrzashova, and R. Zheng, "Diabetes prediction model based on an enhanced deep neural network," *EURASIP J. Wirel. Commun. Netw.*, vol. 2020, 2020. https://doi.org/10.1186/s13638-020-01765-7

[39] K. M. Kuo, P. Talley, Y. H. Kao, and C. H. Huang, "A multi-class classification model for supporting the diagnosis of type II diabetes mellitus," *PeerJ.*, vol. 8, p. e9920, 2020. https://doi.org/10.7717/peerj.9920

[40] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automated detection of mild and multi-class diabetic eye diseases using deep learning," *Health Inf. Sci. Syst.*, vol. 8, 2020. https://doi.org/10.1007/s13755-020-00125-5

[41] M. Rizky, A. Pramuntadi, D. Prastowo, and D. Hardan Gutama, "Implementation of deep neural network method on classification of type 2 diabetes mellitus disease implementasi metode deep neural network pada Klasifikasi Penyakit Diabetes Melitus Tipe 2," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1043–1050, 2024. https://doi.org/10.57152/malcom.v4i3.1279

[42] 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, 2021. https://doi.org/10.1016/j.icte.2021.02.004

[43] C. Fiarni, E. M. Sipayung, and S. Maemunah, "Analysis and prediction of diabetes complication disease using data mining algorithm," *Procedia Computer Science*, vol. 161, pp. 449–457, 2019. https://doi.org/10.1016/j.procs.2019.11.144

[44] 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, 2020. https://doi.org/10.1007/s13755-019-0095-z

[45] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, "Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 235–246, 2020. https://doi.org/10.1109/JBHI.2019.2899218

[46] B. P. Nguyen *et al.*, "Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records," *Comput. Methods Programs Biomed.*, vol. 182, p. 105055, 2019. https://doi.org/10.1016/j.cmpb.2019.105055

[47] P. Bala Manoj Kumar, R. Srinivasa Perumal, R. K. Nadesh, and K. Arivuselvan, "Type 2: Diabetes mellitus prediction using deep neural networks classifier," *International Journal of Cognitive Computing in Engineering*, vol. 1, pp. 55–61, 2020. https://doi.org/10.1016/j.ijcce.2020.10.002

[48] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Applied Computing and Informatics*, vol. 18, no. 1/2, pp. 90–100, 2022. https://doi.org/10.1016/j.aci.2018.12.004

[49] 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, 2022. https://doi.org/10.1007/s13755-021-00168-2

[50] Md. K. Hasan, Md. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, pp. 76516–76531, 2020. https://doi.org/10.1109/ACCESS.2020.2989857

[51] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46, 2021. https://doi.org/10.1016/j.ijcce.2021.01.001

[52] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Comput. Methods Programs Biomed.*, vol. 220, p. 106773, 2022. https://doi.org/10.1016/j.cmpb.2022.106773

[53] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, 2014. https://doi.org/10.1109/TKDE.2013.39

[54] N. Endut, W. M. A. F. W. Hamzah, I. Ismail, M. Kamir Yusof, Y. Abu Baker, and H. Yusoff, "A systematic literature review on multi-label classification based on machine learning algorithms," *TEM Journal*, vol. 11, no. 2, pp. 658–666, 2022. https://doi.org/10.18421/TEM112-20

[55] W. Weng, Y. W. Li, J. H. Liu, S. X. Wu, and C. L. Chen, "Multi-label classification review and opportunities," *Journal of Network Intelligence*, vol. 6, no. 2, pp. 255–275, 2021.

[56] W. Liu, H. Wang, X. Shen, and I. W. Tsang, "The emerging trends of multi-label learning," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7955–7974, 2022. https://doi.org/10.1109/TPAMI.2021.3119334

[57] A. Dagliati *et al.*, "Machine learning methods to predict diabetes complications," *J. Diabetes Sci. Technol.*, vol. 12, no. 2, pp. 295–302, 2018. https://doi.org/10.1177/1932296817706375

[58] A. Pavate and N. Ansari, "Risk prediction of disease complications in type 2 diabetes patients using soft computing techniques," in *2015 Fifth International Conference on Advances in Computing and Communications (ICACC)*, 2015, pp. 371–375. https://doi.org/10.1109/ICACC.2015.61

[59] Z. Xie, O. Nikolayeva, J. Luo, and D. Li, "Building risk prediction models for type 2 diabetes using machine learning techniques," *Prev Chronic Dis*, vol. 16, p. 190109, 2019. https://doi.org/10.5888/pcd16.190109

[60] U.S. Centers for Disease Control and Prevention (CDC), "Behavioral risk factor surveillance system (BRFSS)," 2022. [Online]. Available: https://www.cdc.gov/brfss/

[61] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf Process Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[62] S. Kaul and Y. Kumar, "Artificial intelligence-based learning techniques for diabetes Prediction: Challenges and systematic review," *SN Comput. Sci.*, vol. 1, 2020. https://doi.org/10.1007/s42979-020-00337-2

[63] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: What is it and how does it work?" *Int J Methods Psychiatr Res*, vol. 20, no. 1, pp. 40–49, 2011. https://onlinelibrary.wiley.com/doi/epdf/10.1002/mpr.329

[64] J. Bogatinovski, L. Todorovski, S. Džeroski, and D. Kocev, "Comprehensive comparative study of multi-label classification methods," *Expert Syst. Appl.*, vol. 203, p. 117215, 2022. https://doi.org/10.1016/j.eswa.2022.117215

[65] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Curran Associates, Inc., 2017.

9 APPENDIX

Table A1. Factors used in this study based behavioral factors during 2016–2021 [60]

No	Factors	Question	Value
1	GENHLTH	General Health	1: Excellent, 2: Very good, 3: Good, 4: Fair, 5: Poor
2	MENTHLTH	Number of Days Mental Health Not Good	1: 0–10, 2: 11–20, 3: 21–30
3	HLTHPLN1	Have any health care coverage	1: Yes, 2: No
4	CHECKUP	Length of time since last routine check-up	1: <1 y, 2: 1–2 y, 3: 3–5 y, 4: >5 y, 6: Never
5	EXERANY	Exercise in Past 30 Days	1: Yes, 2: No
6	SMOKER	Computed Smoking Status	1: Current smoker every day, 2: Current smoker some days, 3: Former smoker, 4: Never smoked
7	TOTINDA	Leisure Time Physical Activity Calculated Variable	1: Had physical activity or exercise, 2: No physical activity in past 30 days
8	RFDRHV	Heavy Alcohol Consumption Calculated Variable	1: No; 2: Yes
9	MARITAL	Marital Status	1: Married, 2: Divorced, 3: Widowed, 4: Separated, 5: Never married, 6: Unmarried couple
10	RENTHOM1	Own or Rent Home	1: Own, 2: Rent, 3: Other
11	EMPLOY	Employment Status	1: Employed, 2: Self-employed, 3: No work >1 y, 4: No work <1 y, 5: Homemaker, 6: Student, 7: Retired, 8: Unable to work
12	INCOME2	Income Level	1: <$10 K, 2: $10–$15 K, 3: $15–$20 K, 4: $20–$25 K, 5: $25–$35 K, 6: $35–$50 K, 7: $50–$75 K, 8: >$75 K

(Continued)

Table A1. Factors used in this study based behavioral factors during 2016–2021 [60] *(Continued)*

No	Factors	Question	Value
13	EDUCAG	Computed level of education completed categories	1: Did not graduate high school, 2: Graduated high school, 3: Attended college, 4: Graduated college
14	MSCODE	Metropolitan Status Code	1: Centre city, 2: County, 3: Suburban, 5: not in MSA
15	BLIND	Blind or Difficulty seeing	1: Yes, 2: No
16	DECIDE	Difficulty Concentrating or Remembering	1: Yes, 2: No
17	SEX	Calculated sex variable	1: Male, 2: Female
18	FLSHOT	Flu Shot Calculated Variable	1: Yes, 2: No
19	RACE	Computed Race-Ethnicity grouping	1: White, 2: Black, 3: American Indian or Alaskan Native, 4: Asian, 5: Native Hawaiian or other Pacific Islander, 6: Other race, 7: Multiracial, 8: Hispanic
20	BMICAT	Computed body mass index categories	1: Underweight, 2: Normal weight, 3: Overweight, 4: Obese
21	AGEGYR	Reported age in five-year age categories calculated variable	1: 31 to 40 y, 2: 41–50 y, 3: 51–60 y, 4: 61–70 y, 5: 71–80 y, 6: >81 y
22	DEAF	Are you deaf or do you have serious difficulty hearing?	1: Yes, 2: No
23	DIABETE	(Ever told) you have diabetes	1: Yes, 2: Yes but pregnant, 3: No, 4: Prediabetes
24	PDIABTST	Had a test for high blood sugar or diabetes in the past three years?	1: Yes; 2: Yes, during pregnancy; 3: No
25	PREDIAB1	Ever been told by a doctor or other health professional that you have pre-diabetes or borderline diabetes?	1: Yes; 2: Yes, during pregnancy; 3: No
26	RFBMI	Overweight or obese calculated variable	1: No (Notes: 1200 <= _BMI5 < 2500 (BMI5 has 2 implied decimal places); 2: Yes (Notes: 2500 <= _BMI5 < 9999)
27	CHCKDNY	(Ever told) you have kidney disease?	1: Yes, 2: No
28	CVDCRHD	Ever Diagnosed with Angina or Coronary Heart Disease	1: Yes, 2: No
29	CVDINFR	Ever Diagnosed with Heart Attack	1: Yes, 2: No
30	CVDSTRK	Ever Diagnosed with a Stroke	1: Yes, 2: No
31	CHCOCNC	(Ever told) you had any other types of cancer?	1: Yes, 2: No
32	HAVARTH	Told Have Arthritis	1: Yes, 2: No
33	ADDEPEV	Ever told you had a depressive disorder	1: Yes, 2: No

10 AUTHORS

Nur Rachman Dzakiyullah holds S.Kom from Informatics Engineering in Universitas Islam Indonesia, Yogyakarta while M.Sc and Ph.D from Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. Areas of research interest include Industrial Computing, Operation Research, Modelling and Decision Technology, Data Mining, Artificial Intelligence, and Health Informatics. Currently, working as the Dean of Faculty Computer and Engineering, Universitas Alma Ata, Yogyakarta, Indonesia (E-mail: nurrachmandzakiyullah@almaata.ac.id, P031710017@student.utem.edu.my).

M. A. Burhanuddin is a Professor at the Faculty of Information and Communication Technology (FTMK), Universiti Teknikal Malaysia Melaka (UTeM). He is a member of Majlis Profesor Negara Malaysia. He is a former Deputy Director of UTeM-Melaka RICE Management Center; Director of Industrial and Community Centre; Director of UTeM International Centre; Head of Smart Computing and Business Intelligent Cluster in Advanced Manufacturing Centre; Deputy Dean of Research and postgraduate Studies in FTMK; Head of Department of Industrial computing, FTMK UTeM; Biomedical and Engineering Research Group Leader in UTeM. M. A. Burhanuddin has been active at all levels of the university development in FTMK, UTeM. He was appointed as one of the team members from Malaysia to set up Faculty of Computing and Information Technology, King Abdulaziz University Rabigh, Kingdom of Saudi Arabia. He has a huge experience in managing research grants and supervises many undergraduate and postgraduate students. He has a working experience with corporate sectors, including Intel Technology Sdn. Bhd., Esso Production Incorporations and Rubber Industrial Development Authorities. His teaching and research interests focus on multiple criteria decision-making models, computational modeling, decision support system, optimization techniques, operational research, artificial intelligence, machine learning, soft computing and healthcare informatics areas. He was awarded Bronze Medal in Malaysia Technology Expo 2009 (MTE 2009) with product "Web Based Maintenance Decision Support System for Small and Medium Industries," Silver Medal in UTeM Exhibition Expo 2015 (UTeMEX 2015) with product "My Intelligent Vacation Planner Application," and Bronze Medal in UTeM Exhibition Expo 2017 (UTeMEX 2017) with product "Big Data Analytics: Sentiment Analysis Prototype for Telecommunications Providers in Malaysia."

Raja Rina Raja Ikram is a passionate researcher and educator with expertise in Health Informatics, Software Engineering, and Educational Technology in TVET. She has conducted numerous impactful research projects and presented her work at both local and international conferences. Beyond her research, Dr. Raja Rina actively collaborates with organizations to integrate advanced technologies into technical and vocational education, striving to enhance learning experiences and outcomes for future generations.

Novanto Yudistira is currently employed as a Lecturer and Researcher at the Faculty of Computer Science, Universitas Brawijaya, Indonesia. He obtained his Bachelor of Computer Science degree in informatics engineering from the Institut Teknologi Sepuluh November (ITS) in 2007, then earned MSc degree in computer science from Universiti Teknologi Malaysia (UTM) in 2011, and a Doctor of Engineering (Dr. Eng.) degree in information engineering from the Faculty of Engineering, Hiroshima University, Japan in 2018 under supervision of Prof. Takio Kurita. In 2016, he was involved in research collaboration with the Mathematical Neuro-informatics Group, National Institute of Advanced Industrial Science and Technology (AIST),

Japan. In 2018, he continued his Postdoctoral studies in the field of informatics and data analytics for 2 years in collaboration with Japan's largest scientific research institute, RIKEN, and Osaka University. His current research interests include deep learning in computers, multi-modal computer vision learning, medical informatics, integration of large language (LLM) model and large visual model (LVM), and big data analytics. He is currently active as the founder of the Deep Learning Indonesia group, a community dedicated to developing and studying advancements in Deep Learning algorithms in Indonesia. He has published and reviewed various computer vision and pattern recognition journals like IJCV, IEEE Transactions, IEEE Access, etc. and pattern recognition conferences like ICONIP, SMC, ICPR, etc. Additionally, he is currently conducting various research and community service activities in collaboration with various institutions, companies, and universities in Indonesia.

Muhammad Rifqi Fauzi holds a bachelor's degree in informatics engineering from Universitas Brawijaya, Malang where he specialized in Deep Learning and conducted undergraduate research focused on time series forecasting, aiming to predict future trends based on historical data. Currently, he is pursuing a master's degree in computer science, with a focus on expanding his expertise in Computer Vision, driven by a passion for applying AI to solve complex visual processing challenges.

Dwi Joko Purbohadi is Associate Professor in the Department of Information Technology at Universitas Muhammadiyah Yogyakarta in Indonesia, he deeply immersed in the dynamic realms of e-learning, programming, and mathematics. With a distinguished Ph.D. in Education from Universitas Gadjah Mada, and a rich tapestry of 27 years in academia, his passion lies in pioneering research that reshapes the future of education. He endeavors span the horizon of mobile app development, the ingenuity of intelligence tutoring systems, and the transformative power of technology-enhanced learning environments. Honored to contribute as a reviewer for esteemed academic journals and a keynote presenter at both national and international conferences, his journey continues to be defined by innovation, insight, and impact. His interests include technology, content, and method in e-learning development. He has expertise in qualitative research methods and have published several articles in peer-reviewed journals related to these areas.