Automated Machine Learning for Prediction of Type 2 Diabetes and Its Major Complications: A Comparative Study

Dasun Mallikarachchi
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20166892@my.sliit.lk

Dulakshi Rathnayake
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20142964@my.sliit.lk

Dulmi Abegunawardana
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20141288@my.sliit.lk

Shannon Van-Hoff
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20089290@my.sliit.lk

Dharshana Kasthurirathna
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
dharshana.k@sliit.lk

A Gamage
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
anjalie.g@sliit.lk

2023 5th International Conference on Advancements in Computing (ICAC) | 979-8-3503-5813-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAC60630.2023.10417572

***Abstract*—Type 2 Diabetes (T2D) is a prevalent, chronic health condition linked to complications like Chronic Kidney Disease (CKD) and Ischemic Heart Disease (IHD). Early prediction of T2D and these complications is pivotal for preventive healthcare and treatments. However, implementing machine learning (ML) techniques for such predictions demands specialized expertise, making them inaccessible to non-experts. Additionally, conventional ML methods might offer moderate accuracy and require extensive manual effort during model development. This research introduces an automated machine learning (AutoML) approach that empowers non-experts to train ML models and predict T2D, IHD, and CKD, facilitating early intervention and improved patient outcomes. AutoML automates the entire ML process, from data preprocessing and feature selection to model choice and hyperparameter tuning. To evaluate its effectiveness, benchmark ML models were developed and compared with AutoML models generated from AutoSK-Learn, TPOT and H2O AutoML regarding various performance metrics. The AutoML pipeline models delivered highly accurate predictions for T2D, CKD, and IHD, while significantly r educing d evelopment t ime. Auto-Sklearn and TPOT emerged as best-performing AutoML tools, outperforming the benchmark conventional ML models. This automated pipeline identified e ssential f eatures a nd optimized model hyperparameters, demonstrating robust generalizability and scalability for potential integration into clinical decision support systems.**

Index Terms*—*Automated Machine Learning (AutoML), Auto-Sklearn, TPOT, H2O AutoML, Type 2 Diabetes, Chronic Kidney Disease, Ischemic Heart Disease, Machine Learning

I. Introduction

In recent decades, the prevalence of Type 2 Diabetes Mellitus has escalated dramatically, emerging as a global health challenge with far-reaching consequences. Apart from its direct impact on individual health, Type 2 Diabetes (T2D) is also associated with severe complications; the two most predominant of them are namely Ischemic Heart Disease (Also commonly referred to as Coronary Artery Disease) (IHD) and Chronic Kidney Disease (CKD). Early detection and accurate prediction of these complications are crucial for effective patient management and preventive strategies.

The increasing pervasiveness of T2D is closely tied to modern lifestyle changes, such as sedentary behavior, unhealthy diets, and also genetic factors. This has resulted in a rise in diabetes cases, which in turn has led to higher incidences of IHD and CKD, placing additional strain on global healthcare systems. Traditional risk prediction models have limitations like low accuracy, manual feature and model selection and parameter tuning, and inefficiency in handling large datasets. In contrast, automated machine learning (AutoML) techniques automate model creation, optimize feature selection, and enhance hyperparameter tuning. This yields better prediction accuracy and faster results. The integration of automated machine learning and advanced ML methods has transformed medical research and clinical practices. By harnessing artificial intelligence and vast datasets, researchers and clinicians can develop predictive models for their specific regions and populations, even without extensive ML expertise. This has the potential to greatly improve disease risk assessment and personalized treatment planning. This research paper explores and compares the application of AutoML and ML-based techniques to predict Type 2 diabetes and its two most widespread complications, namely Ischemic Heart Disease (IHD) and Chronic Kidney Disease (CKD). The goal is to evaluate the potential of advanced AutoML tools like Auto-Sklearn, TPOT, and H2O AutoML in creating accurate and accessible predictive models for identifying individuals at risk of these conditions. The study intends to meticulously compare these AutoML pipelines with traditional ML models, focusing on factors like computation time, model complexity, and predictive accuracy.

By synthesizing existing research, real-world case studies, and practical implementations, this study aims to contribute valuable insights into the potential applications of automated machine learning and ML-based approaches in predicting T2D and its complications. Ultimately, our findings aspire to support healthcare professionals in making well-informed decisions, optimizing patient care, and formulating effective preventive strategies to curb the rising tide of T2D and its grave consequences.

II. Related Works

Numerous research studies have focused on using machine learning algorithms to predict T2D and its complications [1] [2]. Researchers have utilized different machine-learning approaches to tackle the predictive challenges of T2D. Moreover, they have also explored the realm of AutoML to improve prediction accuracy and to address the challenges of making machine learning accessible to non-experts [3] [4], allowing them to harness its potential without in-depth data science expertise.

Kumari et al. [1] presented an ensemble ML technique to enhance early-stage diabetes detection. Their method combined random forest, logistic regression, and Naive Bayes algorithms using a soft voting classifier on the Pima Indians Diabetes dataset. The approach achieved top accuracy (79.04%), precision (73.48%), recall (71.45%), and F1-score (80.6%) compared to other methods. They also demonstrated its efficacy in breast cancer prediction, aligning with our work on automated ML for predicting Type 2 diabetes and its complications.

A study carried out by Tigga et al. [2] focused on assessing the risk of T2D among individuals based on lifestyle and family background. Different machine learning algorithms were used for prediction due to their high accuracy, which is crucial in the health profession. The experiment involved 952 instances collected through questionnaires, and the Random Forest Classifier showed the most accurate performance for both the study dataset and the Pima Indian Diabetes database. The findings suggest the potential of machine learning-based approaches for early diabetes risk assessment and self-assessment by individuals.

Literature shows that many Machine learning models have been devised to predict ischemic heart disease. Park et al.(2021) [5] utilized a dataset of 1066 patients with acute ischemic stroke between January 2019 and March 2021 to predict functional outcomes in patients with acute ischemic stroke and ischemic heart disease. Five machine learning algorithms were applied, with regularized logistic regression showing the best performance. Support vector machines also performed well. The National Institute of Health Stroke Scale and age were consistently identified as important variables. These models were validated and found to be pivotal in predicting functional outcomes in acute ischemic stroke.

Prior studies have explored diverse AutoML tools that leverage genetic algorithms, neural architecture search, Bayesian optimization, and meta-learning for disease prediction. [3] [6] [4] [7] These approaches efficiently discover optimal models, hyperparameters, and features, boosting predictive performance.

The scientific review conducted by Waring et al. [3] focuses on AutoML in healthcare, aiming to enable healthcare professionals with limited data science expertise to utilize machine learning models ”off the shelf.” It matches or surpasses human performance in some tasks, though scalability to large datasets remains a challenge. AutoML can bridge expertise gaps in healthcare, selecting and optimizing models for specific tasks.

Padmanabhan et al. [7] applied AutoML for cardiovascular disease prediction. They compared it with manual ML models built by masters students using Scikit-learn. AutoML yielded competitive classifiers with much shorter development time, showcasing similar or superior accuracy and better generalization. AutoML's potential to simplify machine learning for biomedical researchers is underscored, promoting AI adoption in clinical realms.

Based on the literature review performed, it becomes evident that there is a research gap in the domain of utilizing AutoML techniques for the prediction of Type 2 diabetes and its associated complications. This unexplored area offers a promising avenue for further research.

III. Methodology

This study follows a structured methodology involving key stages. Firstly, pertinent datasets for T2D, IHD, and CKD are sourced from reputable publicly available platforms. These datasets form the foundation for predictive modeling. Secondly, data preprocessing and feature selection ensure dataset quality, relevance, and suitability for machine learning algorithms. Following this, individual base machine-learning models are developed for each disease, accounting for specific disease characteristics. These models serve as benchmarks for evaluating conventional ML against AutoML performance. AutoML pipelines, utilizing cutting-edge frameworks, are then constructed to streamline model development. Lastly, an extensive comparative analysis is undertaken, juxtaposing AutoML pipeline models with conventional ML models. This comprehensive assessment employs diverse metrics, facilitating a comprehensive evaluation of their respective performance and capabilities. Figure 1 depicts the methodology flow of this study.

A. Type 2 Diabetes Prediction

We aimed to develop an automated machine learning (AutoML) approach to predict Type 2 Diabetes(T2D) occurrence

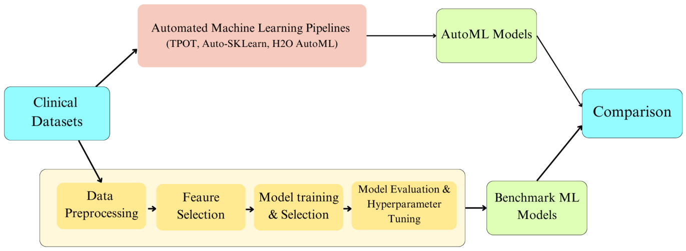

Fig. 1. Proposed Methodology for AutoML vs Conventional ML comparison

using the 2015 US Behavioral Risk Factor Surveillance System (BRFSS) dataset [8], a publicly available dataset that contains results from a survey that collects data on various health-related behaviors, chronic conditions, and use of preventive services among adults. The dataset encompasses around 254,000 instances and 21 features pertaining to occurrences of T2D within the adult population and one binary target class specifying whether the subject has T2D or not. To establish a performance benchmark, ensemble machine learning (ML) models were constructed using various ensembling techniques, including methods such as bagging, boosting, and stacking. The base models that were trained for building these Ensemble models included Support Vector Machine, Decision Tree, Logistic Regression, Random Forest, and K-Nearest Neighbor algorithms.

TABLE I
BENCHMARK ML MODELS TRAINED FOR T2D PREDICTION

Boosting	XGBoost Ensemble
Voting	Soft Voting Ensemble, Hard Voting Ensemble
Bagging	Bagged Soft Voting Ensemble
Stacking	Stacked Ensemble (Metaclassifier: Random Forest)

The feature selection process for constructing the base ML models involved employing univariate feature selection techniques like ANOVA, chi-squared test, and covariance analyses. These aimed to identify significant attributes linked to T2D. Following a rigorous examination, a subset of 16 features with robust associations was thoughtfully curated. Due to the dataset being imbalanced, where the negative class label occurred only in 30% of the records, the Near Miss algorithm was employed to address this disparity. Standard Scaler ensured data normalization, while grid search cross-validation was employed to tune hyperparameters, boosting base ML model accuracy. This method systematically explores parameter combinations within predefined grids, aligning with the overarching objective of enhancing predictive accuracy.

AutoML was then leveraged to accelerate and enhance the modeling process and subsequently test whether it provides increased accuracy for Type 2 Diabetes prediction in comparison to traditional ML models. Two distinct AutoML frameworks, namely Auto-Sklearn, and TPOT, were employed.

Auto-Sklearn [9] is a cutting-edge AutoML approach that integrates and automates dataset cleansing, feature engineering, algorithm selection, and hyperparameter optimization. Operating on a foundation of meta-learning and Bayesian optimization, it refines model development by leveraging historical performance data across diverse datasets. This fusion of techniques expedites model creation while ensuring accuracy. Noteworthy parameters, such as time_left_for_this_task and per_run_time_limit, afford precise control over optimization duration. Adaptive to various data types and tasks, Auto-Sklearn is a versatile tool for streamlined and effective AutoML.

Conversely, TPOT AutoML [10] is also a leading solution in AutoML, adeptly integrating data pre-processing, feature engineering, algorithm selection, and hyperparameter optimization. It employs genetic programming and exhaustive search techniques to accelerate model development and enhance predictive performance. Key parameters like generations, population_size, and max_time_mins, set by users, determine the trade-off between optimization duration and outcome quality. The generations parameter defines the number of iterations TPOT will run to find the optimal pipeline, while population-size parameter influences solution diversity by defining candidate pipelines per generation. TPOT's adaptability across various data types and tasks makes it a crucial tool for efficient AutoML processes. Our experiments systematically explored different permutations of population_size, generations, and max_time_mins parameters to identify optimal values for achieving accurate AutoML model results.

Through the integration of these methodologies, an extensive evaluation of the predictive capabilities of the AutoML pipelines was conducted. By comparing their performance with the previously established ensemble ML models, particularly the XGBoost benchmark, the research aimed to determine whether the automated approaches could yield superior predictive accuracy and streamline the overall model development process. As a result, the methodology encompassed a comprehensive process of data preparation, feature selection, benchmarking, and AutoML integration, culminating in a robust assessment of the AutoML methods' effectiveness in predicting diabetes occurrence.

B. Ischemic Heart Disease Prediction

In the context of predicting Ischemic Heart Disease, the dataset utilized was sourced from a publicly accessible healthcare repository in South Africa, with a focus on Ischemic Heart Disease. This dataset comprises 462 instances, balanced with approximately two controls per disease case. It encompasses 9 distinct feature variables, with the primary target variable indicating the presence or absence of ischemic heart disease.

Data preprocessing techniques were employed, including handling missing values and outliers and data normalization through standardization. Additionally, strategies were applied to rectify data imbalance, ensuring consistent model performance across risk levels. Feature selection utilized robust algorithms, such as Recursive Feature Elimination (RFE), and significance estimation techniques like correlation analysis and tree-based models. The selected features underwent evaluation for clinical relevance and predictive enhancement for ischemic heart disease and ischemic stroke risk assessments.

To maintain risk level distribution integrity, the preprocessed dataset was stratified into distinct training and testing subsets, adhering to an 80-20 split ratio. Standard scaling was then applied for data normalization, enabling equitable algorithmic comparisons.

Benchmark ML models encompassed a diverse array of algorithms, including XGBoost, GBM, SVM, CatBoost, Random Forest, GNB, Decision Tree, Logistic Regression, and K Nearest Neighbor.

AutoML pipelines were constructed using two libraries: TPOT and H2O AutoML. These methods meticulously explore hyperparameter configurations across algorithms like Gradient Boosting, Random Forest, and XGBoost. This comprehensive approach sought optimal model configurations, iteratively experimenting with hyperparameter combinations. Performance metrics like accuracy and F1-score objectively evaluated and compared model performances.

C. Chronic Kidney Disease Prediction

The objective of this section was to create an AutoML approach to predict Chronic Kidney Disease (CKD) and compare its performance with conventional ML models. The CKD dataset for this study, sourced from the UCI Machine Learning Repository [11], consists of 400 samples with 24 features and a binary class label: ”ckd” or ”notckd.”

Preprocessing involved addressing outliers, noise, and missing values. Imputation methods estimated missing data—mode for nominal features and mean for numerical ones. The dataset was then divided into 70% training and 30% testing subsets, normalized using Standard Scaler.

ML algorithms such as Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), CatBoost, and Gradient Boosting—formed the base ML models, and the best performing one was to be chosen as the benchmark ML model for the comparison.

AutoML frameworks TPOT and H2O.ai AutoML constructed automated pipelines. These pipelines automated dataset cleansing, feature engineering, and model selection, training, optimization, and evaluation, thus having the ability to predict CKD likelihood for new patients using new datasets in practical healthcare scenarios.

IV. RESULTS AND DISCUSSION

A. Type 2 Diabetes Prediction

To evaluate the performance of the benchmark ensemble ML models that were trained and their base ML models, the metric of accuracy was employed. It was observed that the XGBoost ensemble algorithm achieved the highest performance, outperforming other ensemble and base models with an accuracy of 0.866 and an MSE value of 0.133672. Table 2 compares the accuracy and MSE values of the trained ML models.

TABLE II
T2D BENCHMARK ML MODEL PERFORMANCE METRICS

Model Name	Accuracy
XGBoost	0.866328
Bagged Soft Ensemble	0.865283
Stacked Ensemble	0.864334
SVM	0.863526
Soft Ensemble	0.863146
Random Forest	0.863051
Hard Ensemble	0.860725
Decision Tree	0.850610
Logistic Regression	0.847666
K Nearest Neighbor	0.810627

To evaluate the AutoML pipelines that were exercised in this study, Accuracy, precision, and recall were utilized. In the context of the TPOT AutoML pipeline, extensive experimentation was conducted with the TPOTClassifier with different permutations of hyperparameters. The hyperparameter configuration, which involved setting the generations parameter to 30, population size to 100, and a substantial maximum time limit of 480 minutes (equivalent to 8 hours), led to the identification of the Random Forest Classifier with n_estimators parameter as 100 as the best-performing pipeline. This pipeline achieved an accuracy score of 8.675. Notably, it outperformed all established benchmark ensemble ML models. The empirical observation of the TPOT AutoML pipeline underscores a noteworthy aspect: the pipeline's effectiveness improves with extended runtime,

The utilization of the Auto-Sklearn automl pipeline involved running the pipeline with various permutations of hyperparameter configurations. The pipeline configured with ”time_left_for_this_task” set at 10800 seconds (3 hours) and ”per_run_time_limit” at 600 (10 minutes), yielded promising results. Notably, the achieved accuracy score of 0.868 surpassed the performance of all previously established benchmark ensemble machine learning models and also the performance of TPOT best pipeline. This outcome underscores the effectiveness of the Auto-Sklearn approach in delivering superior predictive accuracy within the specified time constraints, ultimately yielding superior results. The figures 2 and 3 exemplify the high predictive accuracy of Auto-Sklearn.

B. Ischemic Heart Disease Prediction

The investigation revealed that among the array of implemented ML models aimed at selecting an optimal benchmark for Ischemic Heart Disease (IHD) prediction, the Support

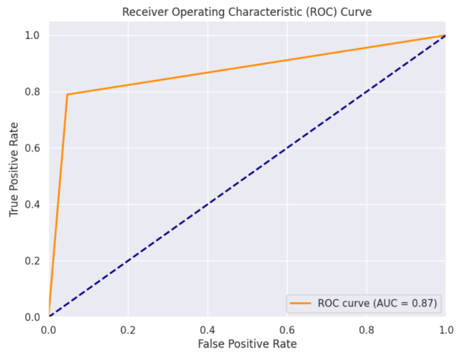

Fig. 2. AUC-ROC curve of Auto-Sklearn best pipeline for T2D Prediction

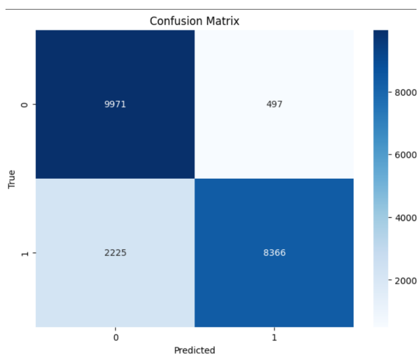

Fig. 3. Confusion Matrix of Auto-Sklearn best pipeline for T2D Prediction

Vector Machine (SVM) model exhibited superior performance. It notably surpassed other models in metrics encompassing accuracy, precision, recall, and F1-score. The SVM model demonstrated an accuracy level of 0.732 with an accompanying F1-score of 0.79. Table III compares the predictive performances of the two trained benchmark ML models.

Regarding AutoML pipelines, the TPOT AutoML pipeline yielded an ExtraTreesClassifier model as the best-performing pipeline. This model was characterized by the following specific hyperparameter settings: (input_matrix, bootstrap=False, criterion=gini, max_features=0.8, min_samples_leaf=20, min_samples_split=12, n_estimators=100). It was impressive to witness the autoML pipeline achieving an accuracy of 0.97368, thereby surpassing the highest-performing benchmark machine learning model (SVM) by nearly 20% in terms of accuracy. Table IV compares the predictive performances of the two trained AutoML models.

Conversely, the H2O AutoML framework showed commendable performance, surpassing the benchmark SVM model With an accuracy of 0.7456. Intriguingly, its highest-performing pipeline comprised an XGBoost model. This outcome underscores the framework's effectiveness in identifying intricate model configurations that lead to enhanced predictive capabilities.

TABLE III
IHD PREDICTION BENCHMARK MODEL PERFORMANCES

Model Name	Accuracy	Precision	recall	F1 - score
SVM	0.732758	0.72	0.79	0.76
Logistic Regression	0.724137	0.76	0.84	0.79
GNB	0.698275	0.74	0.70	0.72
CatBoost	0.689655	0.73	0.82	0.77
GBM	0.681034	0.70	0.77	0.73
XGBoost	0.672413	0.68	0.71	0.70
K Nearest Neighbor	0.655172	0.66	0.70	0.68
Random Forest	0.646551	0.71	0.78	0.75
Decision Tree	0.620689	0.71	0.68	0.69

TABLE IV
IHD PREDICTION AUTOML PIPELINE PERFORMANCES

AutoML	Accuracy	Precision	Recall	F1 -score
TPot	0.97368	0.9722	0.9859	0.9790
H2O (XGBoost)	0.7456	0.70	0.60	0.61

C. Chronic Kidney Disease Prediction

Through the analysis, it was determined that among the employed machine learning models aimed at establishing an optimal benchmark for the prediction of CKD, the Random Forest model emerged as the most superior performer with an accuracy score of 0.99167. It notably outperformed other models in terms of accuracy.

TABLE V
CKD PREDICTION BENCHMARK ML MODEL PERFORMANCES

Model Name	Accuracy	Precision	recall	F1-Score
Random Forest	0.99167	0.9777	1.0	0.9887
Logistic Regression	0.99167	0.9777	1.0	0.9887
XGBoost Classifier	0.98333	0.9777	1.0	0.9887
Support Vector Machine	0.98333	0.9777	1.0	0.9887
CatBoost	0.98333	0.9777	1.0	0.9887
GBM	0.98333	0.9777	1.0	0.9887
K Nearest Neighbor	0.96667	0.8979	1.0	0.9462
Decision Tree	0.95834	1.0	0.9772	0.9885

Moving into the automated machine learning (AutoML) pipelines, the TPOT AutoML pipeline selected a Gaussian Naive Bayes Classifier as its top-performing model. This constructed pipeline model achieved an accuracy of 0.9667. This considerably modest accuracy might be due to the lower max_time_mins hyperparameter set during the TPOT experimentation, and it is anticipated that enhancing this value could lead to improved accuracy in the outcomes.

On a different note, the H2O AutoML framework demonstrated high effectiveness, outperforming both the benchmark Random Forest model and the TPOT AutoML pipeline. With

TABLE VI
CKD PREDICTION AUTOML PERFORMANCES

AutoML	Accuracy	Precision	recall	F1-Score
Tpot	0.9667	0.8918	1	0.9428
H2O (XGBoost)	0.99646	0.9736	1	0.9866

an accuracy of 0.99646, the H2O AutoML pipeline also showed superiority over the benchmark machine learning algorithms. It's worth mentioning that the best-performing pipeline configuration, in this case, was an XGBoost model.

CONCLUSION AND FUTURE WORK

In this study, we conducted an in-depth comparison between conventional ML approaches and AutoML techniques for predicting Type 2 Diabetes and its two most predominant complications: Ischemic Heart Disease and Chronic Kidney Disease. We comparatively evaluated state-of-the-art AutoML pipeline frameworks and conventional benchmark machine learning models using various performance metrics, providing insights into their predictive capabilities.

Our discoveries indicate that employing AutoML techniques through frameworks like TPOT, Auto-Sklearn, and H2O AutoML yielded superior predictive performance compared to conventional machine learning models for forecasting T2D, CKD, and IHD. Auto-Sklearn produced the best-performing model for T2D prediction (Accuracy = 0.868), while TPOT delivered the highest accuracy models for both CKD (0.99646) and IHD (0.7456). Moreover, these AutoML approaches demanded considerably less time, manual work investment, and extensive ML expertise for model development - this implies that medical professionals can train ML models with ease without extensive expertise in the field using AutoML. The efficacy and accuracy of the AutoML pipeline models exhibited a linear relationship with the time duration allocated for their pipeline execution. However, due to constraints imposed by time limitations, the execution durations of AutoML pipelines were restricted to a maximum of 8 hours where needed.

While this study sheds light on the comparative performance of conventional machine learning and AutoML techniques for the prediction of T2D, CKD, and IHD, there remain avenues for future exploration and enhancement in this field, such as the ones listed below:

- Algorithm Tuning and Optimization: Further refinement of the AutoML pipelines could involve extensive hyperparameter tuning, with more hyperparameter permutations and increased allowed run time for the pipelines, potentially leading to even more robust predictive performance.
- Model Interpretability: Enhancing the interpretability of the complex models generated by AutoML frameworks is a crucial direction. Developing techniques to explain the decision-making process of these models will aid in building trust and understanding in medical applications for clinicians and medical professionals.
- Transferability and Generalization: Evaluating the generalization capabilities of the AutoML pipelines on diverse datasets is pivotal. This approach empowers clinicians and medical professionals globally, enabling them to utilize regional data for training AutoML models without the necessity of extensive ML expertise.
- Longitudinal Studies: Expanding the scope of this research to encompass longitudinal data could provide insights into disease progression prediction using AutoML, allowing for early interventions and personalized treatment strategies.

In conclusion, this study underscores the transformative potential of AutoML for predictive analytics in healthcare. The demonstrated superiority of AutoML pipelines over traditional approaches, along with the nuances revealed through our analysis, opens doors for future research that can lead to more accurate and robust predictive models in the realm of Type 2 Diabetes, CKD, IHD, and also other medical conditions.

REFERENCES

[1] 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*, 2:40–46, June 2021.
[2] N. P. Tigga and S. Garg. Prediction of type 2 diabetes using machine learning classification methods. *Procedia Computer Science*, 167:706–716, 2020.
[3] Waring J, Lindvall C, and Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif Intell Med*, 104:101822, April 2020.
[4] Y. Rimal, S. Paudel, N. Sharma, and A. Alsadoon. Machine learning model matters its accuracy: a comparative study of ensemble learning and automl using heart disease prediction. *Multimedia Tools and Applications*, pages 1–18, 2023.
[5] Park D, Jeong E, Kim H, Pyun HW, Kim H, Choi YJ, Kim Y, Jin S, Hong D, Lee DW, Lee SY, and Kim MC. Machine learning-based three-month outcome prediction in acute ischemic stroke: A single cerebrovascular-specialty hospital study in south korea. *Diagnostics (Basel)*, 11(10):1909, 2021.
[6] A. Romero, R. A., Y. Deypalan, M. N. Mehrotra, S. Jungao, J. T. Sheils, N. E. Manduchi, and J. H. Moore. Benchmarking automl frameworks for disease prediction using medical claims. *BioData Mining*, 15(1):15, 2022.
[7] M. Padmanabhan, P. Yuan, G. Chada, and H. V. Nguyen. Physician-friendly machine learning: A case study with cardiovascular disease risk prediction. *JCM*, 8(7):1050, July 2019.
[8] Centers for Disease Control and Prevention (CDC). Behavioral risk factor surveillance system (brfss) 2015 dataset. https://www.cdc.gov/brfss/brfssprevalence/index.html, 2015.
[9] Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Tobias Springenberg, Manuel Blum, and Frank Hutter. Auto-sklearn: Efficient and robust automated machine learning. In *Automated Machine Learning*, 2019.
[10] Randal Olson and Jason Moore. Tpot: A tree-based pipeline optimization tool for automating machine learning. In *Automated Machine Learning*, 2019.
[11] Markelle Kelly, Rachel Longjohn, and Kolby Nottingham. The uci machine learning repository. https://archive.ics.uci.edu.