ICAICTA
International Conference on Advanced Informatics:
Concepts, Theory and Applications

SUPERVISED LEARNING MODELS FOR EARLY DETECTION OF ALBUMINURIA RISK IN TYPE-2 DIABETES MELLITUS PATIENTS

1st Arief Purnama Muharram; 2nd Dicky Levenus Tahapary;
3rd Yeni Dwi Lestari; 4th Randy Sarayar;
5th Valerie Josephine Dirjayanto

1

INTRODUCTION



BACKGROUND

- **Diabetes** continues to be one of the most challenging non-communicable diseases worldwide.
- The predominant challenge associated with diabetes stems from the array of complications that can arise when diabetes is not adequately controlled. Among these unwanted complications, one particularly notable issue is **kidney complication.**
- Diagnosing **diabetic nephropathy** involves considering various criteria, one of which is the presence of a pathologically significant quantity of albumin in urine, known as **albuminuria**.
- **Albuminuria** can signal kidney problems, and it is essential for patients with diabetes to be aware of their risk of developing this condition.
- Early prediction of albuminuria in diabetic patients holds the **potential for timely preventive measures.**

International Diabetes Federation, 2021
KDIGO, 2012
Pavkov et al, 2018

OBJECTIVES

- **Primary**
  - To develop a supervised learning model capable of predicting the risk of albuminuria development in diabetes patients, particularly those with T2DM
- **Specified**
  - To develop a supervised model capable of predicting early albuminuria in patients with type 2 diabetes mellitus (T2DM)
  - To identify optimal supervised algorithm for early albuminuria detection in T2DM patients

2

RELATED WORK



PREVIOUS STUDIES

- **Khitan et al. (2021)**
  - Machine learning approaches to predict the risk of albuminuria in 1330 patients with diabetes
  - Variety of algorithms: random forest, gradient boost, logistic regression, support vector machines, multilayer perceptron, and a stacking classifier.
  - 13 predictive factors, including measures such as subtotal lean mass, subtotal fat mass, diabetes duration, age, HbA1c levels, creatinine levels, triglyceride levels, total cholesterol levels, HDL cholesterol levels, maximum exercise capacity, systolic and diastolic blood pressure, and ankle brachial index (ABI).
  - Multilayer perceptron (MLP) exhibited the highest performance with an AUC (Area Under the Curve) value of 0.67, precision of 0.61, recall of 0.67, and accuracy of 0.62

PREVIOUS STUDIES

- **Lin et al. (2022)**
  - Machine learning approaches to predict the risk of albuminuria in 3,294 Chinese diabetic patients using machine learning approaches
  - Using ”glm” package in the R software
  - Systolic and diastolic blood pressure, fasting blood glucose levels, triglyceride levels, gender, age, and smoking
  - Specificity of 0.9, accuracy of 0.63, sensitivity of 0.2

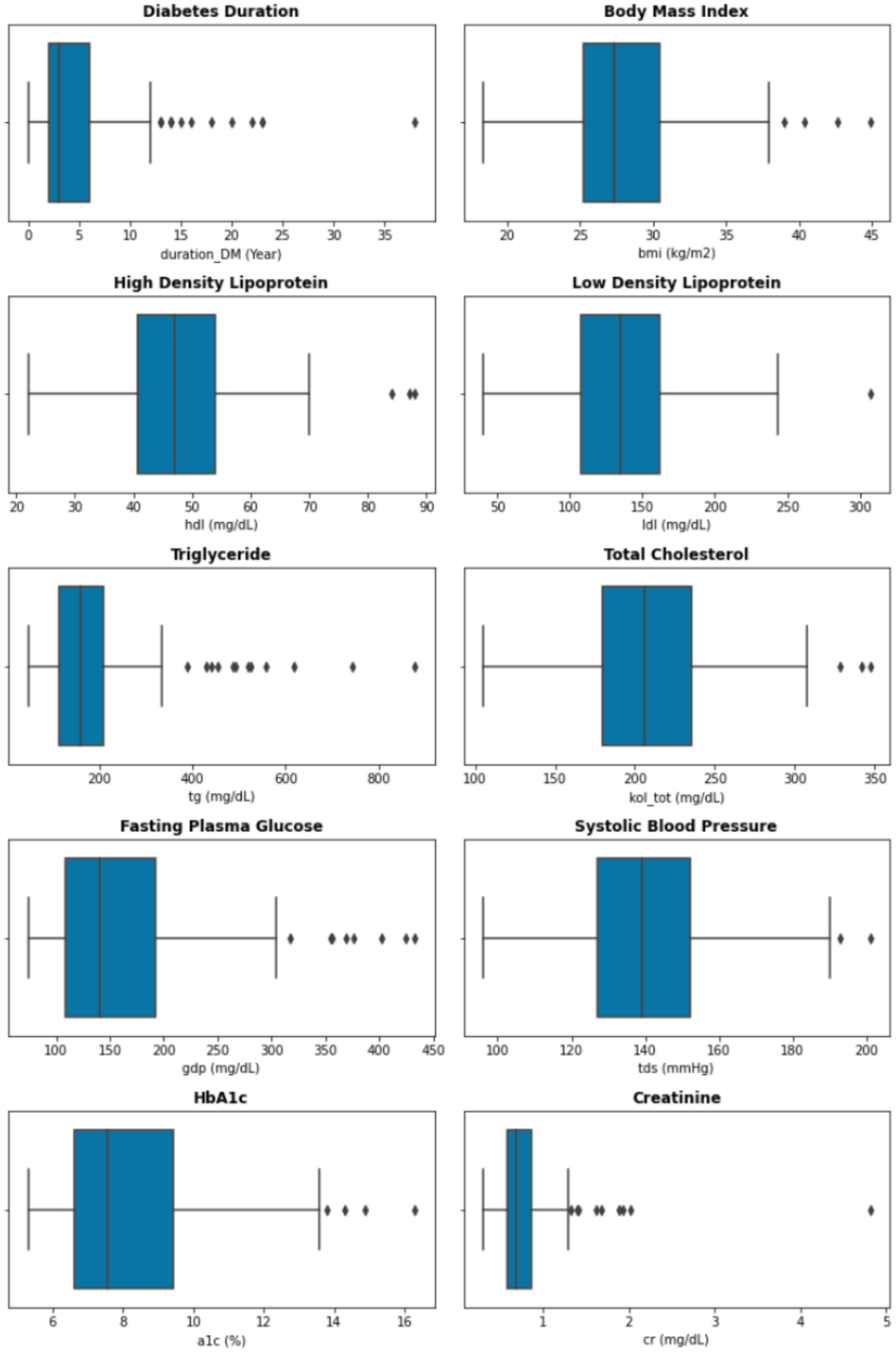
3

DATASET



DATASET

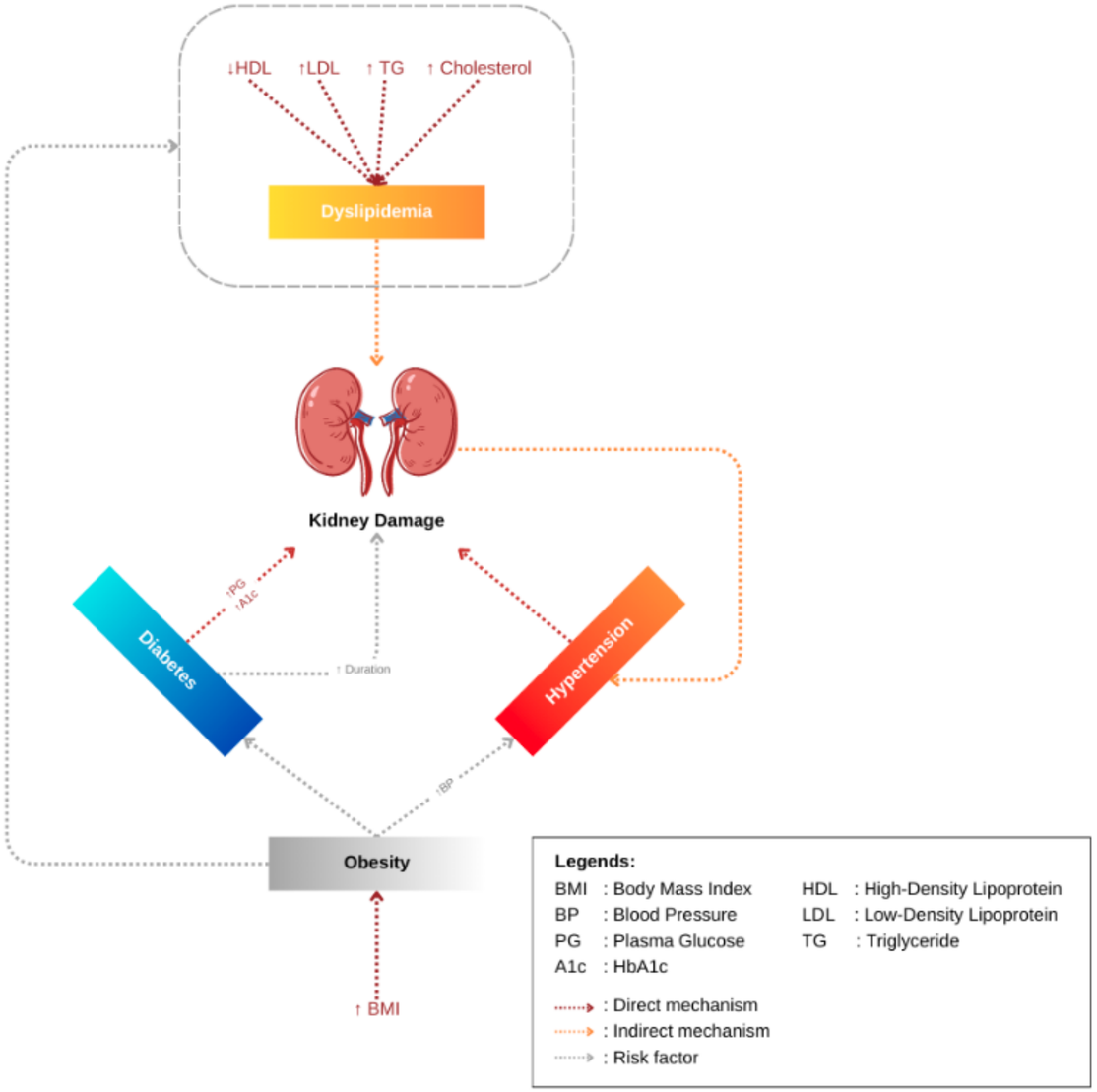
- Anonymized private dataset of 184 records, each consisting of 10 features and 1 target variable



SUMMARY STATISTICS

Attribute	Unit	Mean±Std	Min	Max
durasi_dm	Year	5.168±5.012	0.000	38.000
bmi	kg/m^2	27.973±4.517	18.350	44.870
hdl	mg/dL	48.065±10.913	22.000	88.000
ldl	mg/dL	135.978±38.522	40.000	307.000
tg	mg/dL	185.571±118.305	47.000	878.000
kol_tot	mg/dL	207.554±42.534	105.000	347.000
gdp	mg/dL	158.076±68.591	74.000	433.000
TDS	mmHg	141.609±18.682	96.000	201.000
a1c	%	8.204±2.134	5.300	16.300
cr	mg/dL	0.782±0.409	0.300	4.810
Attribute	**Category**		**Count**	**Perc.**
kid_group	Normal (0)		92	0.50
	Albuminuria (1)		92	0.50

SIMPLIFIED MECHANISM OF DIABETIC NEPHROPATHY



EXPERIMENT DESIGN

- **Scikit-learn library 1.0.2** - primary machine and deep learning (MLP)
- **XGBoost 1.6.2** (for XGBoost)
- Train-test ratio: 0.75:0.25

Algorithm	Observable Factors	
Naïve Bayes	**Class**	sklearn.naive_bayes.GaussianNB
	Params	priors=None, var_smoothing=1e-09
SVM	**Class**	sklearn.svm.SVC
	Params	C=1.0, kernel='rbf', max_iter=-1
Decision Tree	**Class**	sklearn.tree.DecisionTreeClassifier
	Params	criterion='gini', max_depth=None
Random Forest	**Class**	sklearn.ensemble.RandomForestClassifier
	Params	n_estimators=100, criterion='gini', max_depth=None
AdaBoost	**Class**	sklearn.ensemble.AdaBoostClassifier
	Params	n_estimators=50, learning_rate=1.0, algorithm='SAMME.R'
XGBoost	**Class**	xgboost.XGBClassifier
	Params	n_estimators=2, max_depth=1, learning_rate=1
MLP	**Class**	sklearn.neural_network.MLPClassifier
	Params	hidden_layer_sizes=(100,), learning_rate_init=3e-3, max_iter=200

EVALUATION STRATEGY

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$F1\text{-}score = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

ETHICS

KET-246/UN2.F1/ETIK/PPM.00.02/2022

4

RESULTS



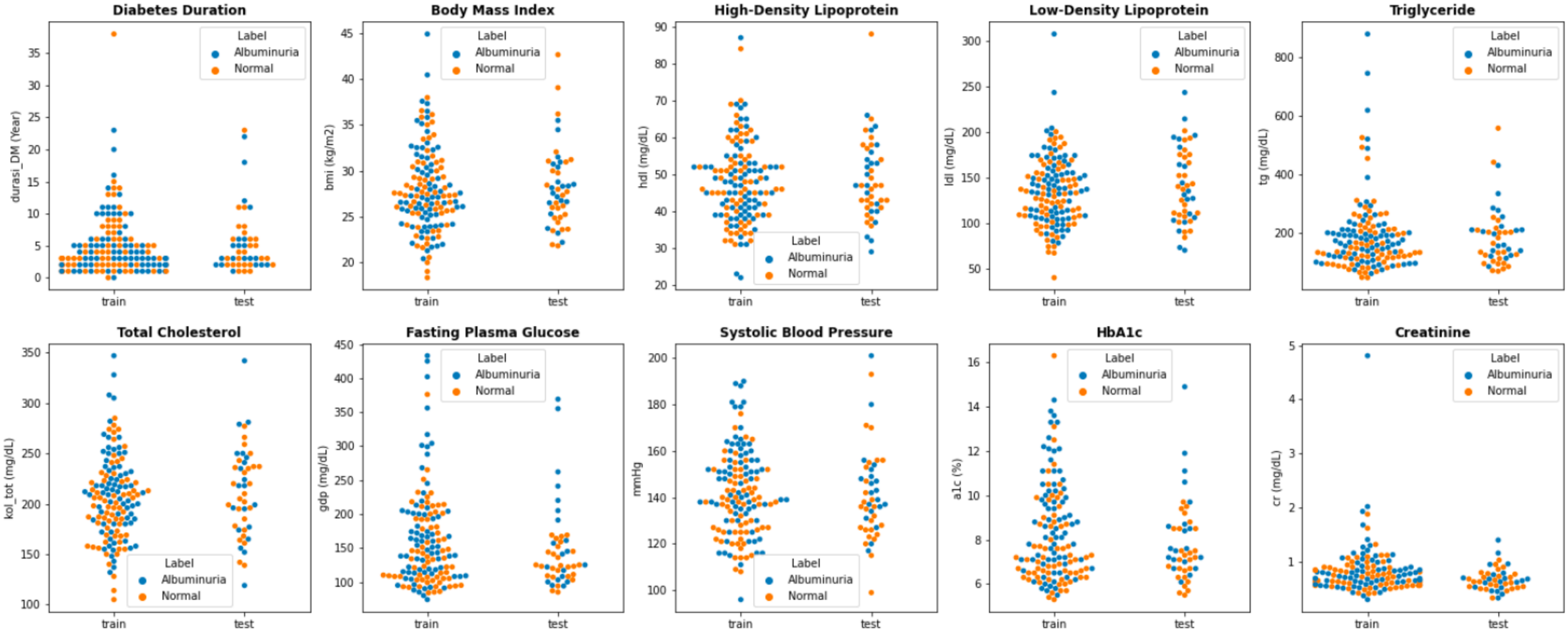
MODEL EVALUATION

- Test ratio: 0.25 -> 138 records for train dataset, 46 for test dataset
- Relatively uniform distribution

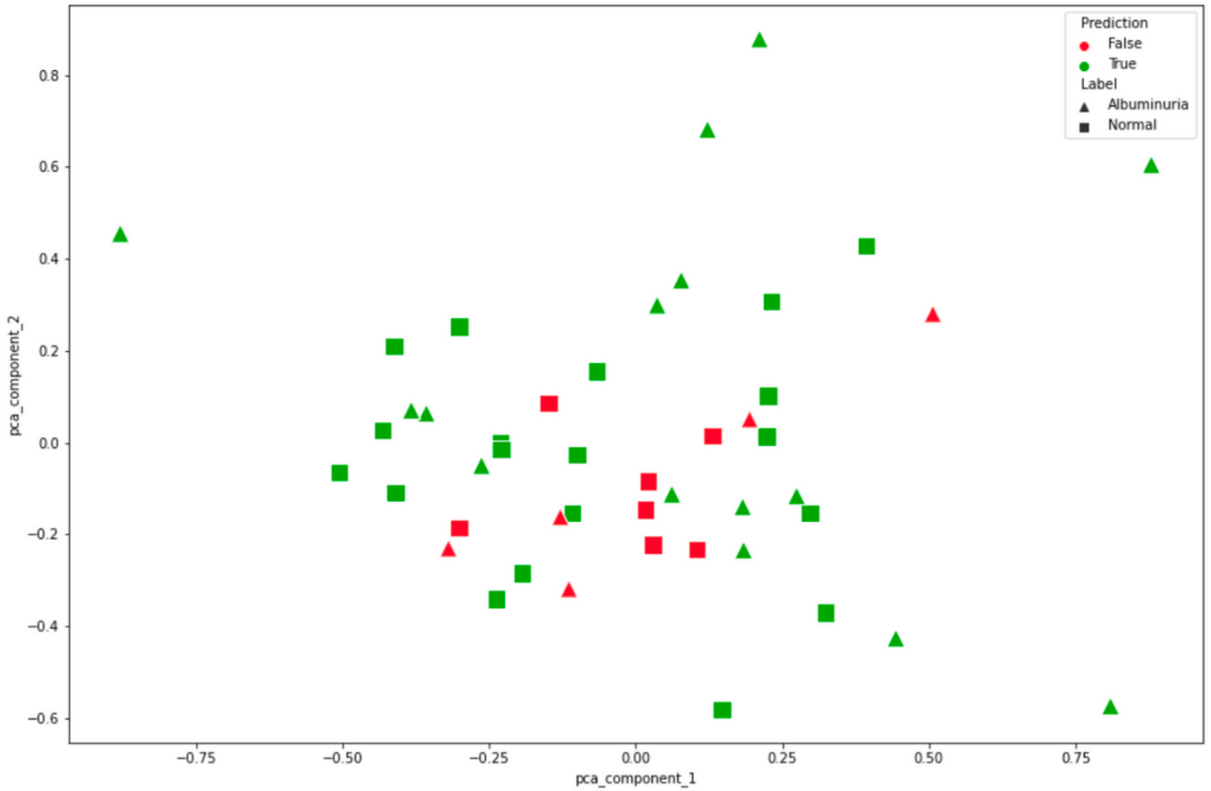
Algorithm	Precision	Recall	Accuracy	F1-Score
Naïve Bayes	0.67	0.4	0.65	0.50
SVM	0.50	0.70	0.57	0.58
Decision Tree	0.45	0.50	0.52	0.48
Random Forest	0.50	0.60	0.57	0.55
AdaBoost	0.44	0.40	0.52	0.42
XGBoost	0.46	0.55	0.52	0.50
MLP	**0.68**	**0.75**	**0.74**	**0.71**

- None of ML algorithm achieved f1-score >0.5
- Superior results from deep learning **MLP**

PRINCIPAL COMPONENT ANALYSIS



ERROR ANALYSIS



- Falsely predicted labels are spread out but relatively close to the adjacent cluster that forms the true predictions
- For several patients, the risk of developing albuminuria might not be strongly correlated with the features in the dataset

STRENGTHS AND LIMITATIONS

- Performance of the MLP algorithm might be beneficial due to its architecture
  - backpropagation algorithm enables the MLP to solve complex problems relatively easily compared to other traditional machine learning algorithms when dealing with complex data

- Should improve the model's accuracy is to increase the dataset size and variety, allowing the learning algorithms to better understand the complex patterns present in such data

5

CONCLUSION



CONCLUSION

- We have developed a supervised learning model to predict the risk of developing albuminuria in patients with T2DM
- Among the various supervised learning models examined in this study, the MLP algorithm demonstrated superior performance in terms of precision, recall, accuracy, and f1-score
- We recommend augmenting the dataset with additional data to increase its size and diversity
- Further research should explore the utilization of deep learning algorithms like MLP to effectively handle the complexities inherent in patient data.

REFERENCES

1. International Diabetes Federation, IDF Atlas. Brussels, Belgium: Inter- national Diabetes Federation, 2021.
2. KDIGO."2012clinicalpracticeguidelinefortheevaluationandmanage- ment of chronic kidney disease," in Kidney International Supplements, vol. 3, pp. 1-150, 2013.
3. M. E. Pavkov, A. J. Collins, J. Coresh, et al., "Kidney Disease in Diabetes," in Diabetes in America, 3rd edition, C. C. Cowie, S. S. Casagrande, A. Menke, et al., Eds., USA: National Institute of Diabetes and Digestive and Kidney Diseases, August 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK568002/. Accessed: July 24, 2023
4. Z. Khitan, T. Nath, P Santhanam, "Machine Learning Approach to Predicting Albuminuria in Persons with Type 2 Diabetes: An Analysis of the LOOK AHEAD Cohort," in The Journal of Clinical Hypertension, vol. 23, no. 12, pp. 2137-2145, 2021. doi: 10.1111/jch.14397.
5. W. Lin, et al., "Development of a Risk Model for Predicting Microalbu- minuria in the Chinese Population Using Machine Learning Algorithms," in Frontiers in Medicine, vol. 9, 2022. doi: 10.3389/fmed.2022.775275
6. O.Z. Ameer, "Hypertension in Chronic Kidney Disease: What Lies Behind the Scene," Frontiers in Pharmacology, vol. 13, p. 949260, Oct. 11, 2022. doi: 10.3389/fphar.2022.949260.
7. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E´. Duchesnay, "Scikit-learn: Machine Learning in Python," in Journal of Machine Learning Research, vol. 12, no. 85, pp. 2825-2830, 2011.
8. T.ChenandC.Guestrin,"XGBoost:AScalableTreeBoostingSystem," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, California, USA, 2016, pp. 785-794. doi: 10.1145/2939672.2939785.
9. T. Mitchell, Machine Learning. USA: McGraw-Hill, 1997.