The Scientific Temper (2024) Vol. 15 (3): 2734-2739
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2024.15.3.42
https://scientifictemper.com/

RESEARCH ARTICLE

Hybridization of bio-inspired algorithms with machine learning models for predicting the risk of type 2 diabetes mellitus

Raja S.*, Nagarajan L.

Abstract

Type 2 diabetes mellitus (T2DM) is a chronic condition that affects millions of people worldwide. Predicting the risk of developing this disease is critical for early intervention and prevention. Bio-inspired algorithms and machine learning models have shown promising results in predicting the risk of T2DM. In this paper, we will explore the use of these two approaches and their hybridization to improve the accuracy of risk prediction. The first section will introduce bio-inspired algorithms and their application in predicting the risk of T2DM. We will discuss the advantages of using these algorithms and their limitations. The second section will focus on machine learning models and their potential to predict the risk of T2DM. We will also discuss the limitations of this approach. The final section will compare and contrast the two approaches and explore how their hybridization can overcome their limitations and improve the accuracy of risk prediction. Overall, this paper aims to provide an in-depth analysis of the use of bio-inspired algorithms and machine learning models in predicting the risk of T2DM and their hybridization to improve their accuracy.

Keywords: Type 2 diabetes mellitus, Bio-inspired algorithms, Machine learning models.

Introduction

Bio-inspired algorithms are innovative techniques inspired by the collective behavior of social colonies, such as ants and bees. These algorithms are proving to be an efficient method for solving complex optimization problems in various fields, including healthcare. One such application is predicting the risk of type 2 diabetes mellitus (T2DM). Bio-inspired algorithms have emerged as an effective problem-solving tool for identifying diabetes disease risks, and they have been found to be quicker and more efficient than traditional methods of diagnosis. Support vector machines and Naïve Bayes algorithms can be employed in conjunction with swarm optimization techniques for the analysis of patterns found in data to diagnose the disease. The healthcare sector is increasingly adopting these techniques to handle complex optimization problems. As such, bio-inspired algorithms are an effective method to predict T2DM and hold great promise for the future of healthcare technology. The hybridization of bio-inspired algorithms with machine learning models has been shown to improve the accuracy of risk prediction. This is achieved by training machine learning models on a set of attributes selected with a bio-inspired evolutionary algorithm. The use of bio-inspired algorithms for optimization purposes has been found to enhance the accuracy of detection of cardiac disease, and there is potential for further improvement in risk prediction accuracy through hybridization with machine learning models. Future work should consider a larger collection of data from a diverse range of subjects for better modeling of classifier optimization for a global population, which could lead to further applications in the field of machine learning. The proposed bio-inspired process includes biological inspiration in every step of the process, resulting in improved accuracy of risk prediction. This process involves the generation of an initial dataset of biological data, selection of attributes *via* biologically inspired computing, optimization of a neural network *via* biologically inspired computing, and use of an optimized neural network for the classification of data. Feature selection algorithms have been applied in medical domain applications, such as diagnosing Parkinson's

PG & Research Department of Computer Science Adaikalamatha College Vallam,Thanjavur Affiliated to Bharathidasan University Tiruchirappalli, Tamil Nadu, India

***Corresponding Author:** Raja S., PG & Research Department of Computer Science Adaikalamatha College Vallam,Thanjavur Affiliated to Bharathidasan University Tiruchirappalli, Tamil Nadu, India, E-Mail: sk.rajamecse@gmail.com

How to cite this article: Raja, S., Nagarajan, L. (2024). Hybridization of bio-inspired algorithms with machine learning models for predicting the risk of type 2 diabetes mellitus. The Scientific Temper, **15**(3):2734-2739.

Doi: 10.58414/SCIENTIFICTEMPER.2024.15.3.42

Source of support: Nil

Conflict of interest: None.

Received: 15/07/2024 **Accepted:** 13/08/2024 **Published** : 16/09/2024

and cancer, with the CFA being used to diagnose Parkinson's at an early stage with an accuracy rate of 94% using KNN. Furthermore, the proposed model outperforms other algorithms in terms of accuracy. Overall, the hybridization of bio-inspired algorithms with machine learning models offers a promising avenue for improving risk prediction accuracy. Bio-inspired algorithms represent a newer field of research that has shown promise in predicting the risk of T2DM. One study utilized machine learning classification algorithms such as logistic and K-nearest to predict the onset of type 2 diabetes. Another study aimed to develop a machine learning model to predict the occurrence of type 2 diabetes in the following year using data from electronic health records. Bio-inspired algorithms offer many advantages, such as incorporating subfields related to connectionism, social behavior and other topics into the prediction process, which can improve the accuracy of predictions. Furthermore, a paper investigated the feasibility of using wearable IoT devices to predict glycemic levels in individuals with type 1 diabetes, indicating that these devices could be used for early risk detection. This shows that bio-inspired algorithms can be used in conjunction with technology to provide more accurate and efficient diabetes detection. Different machine learning techniques have also been proposed to classify a wide range of biological and clinical data, paving the way for more personalized and precise risk prediction models. Overall, bio-inspired algorithms offer a promising avenue for developing more accurate and efficient diabetes risk prediction models.

Existing Machine learning models for predicting the risk of T2DM

Machine learning models have emerged as a useful technique for predicting the risk of T2DM. These models use various features such as age, body mass index, blood pressure, and blood glucose levels to predict the likelihood of developing diabetes. The presented model uses a machine learning technique and is designed to predict the risk of type 2 diabetes. Machine learning algorithms can analyze large datasets to identify patterns and predict outcomes, making them useful for identifying early signs of diabetes. Data exploration through risk factor analysis can help identify associations between features and diabetes. Machine learning techniques can be used to identify individuals at risk of diabetes based on specific risk factors, such as serum electrolytes and physical measurements. The performance of machine learning models in predicting diabetes can be evaluated using metrics such as sensitivity, specificity, accuracy, and area under the curve. Additionally, data pre-processing is a major step in the design of efficient and accurate models for diabetes occurrence. The aim of the systematic review is to identify areas for improving the prediction of type 2 diabetes using machine learning techniques. Random forest, extreme gradient boost, logistic regression, and weighted ensemble model algorithms were employed to predict uncontrolled diabetes, with the random forest model having the highest accuracy in predicting uncontrolled diabetes based on serum variables. Potassium levels, body weight, aspartate aminotransferase, height, and heart rate were important predictors of uncontrolled diabetes. The systematic review analyzed 90 studies to identify the main opportunity areas in diabetes prediction using machine learning. Machine learning models can be used to create predictive models for type 2 diabetes. Complementary techniques can be used to improve the performance of machine learning models in predicting diabetes. The review aims to find areas of opportunity and recommendations in the prediction of diabetes based on machine learning models, and inform the selection of machine learning techniques and features for creating novel type 2 diabetes predictive models.

The use of machine learning models in predicting the risk of T2DM has gained popularity in the healthcare sector due to their ability to handle complex optimization problems. However, there are several limitations associated with using these models. One of the main limitations is the lack of evidence synthesis of the performance of machine learning prediction models for type 2 diabetes. Moreover, the performance of different machine learning and statistical techniques in predicting the risk of T2DM varies. As the number of variables increases, the hypothesis testing method using machine learning models becomes complicated. Additionally, machine learning models can oversimplify complex relationships between nonlinear interaction factors, potentially leading to loss of related information. Furthermore, machine learning models are limited to predicting type 2 diabetes and not other types of diabetes, treatments, or diseases associated with diabetes. Studies using unsupervised learning are excluded from the prediction of type 2 diabetes risk as they cannot be validated using the same performance metrics as supervised learning models. The authors evaluated the technique used, the temporality of prediction, the risk of bias, and validation metrics; however, the best machine learning model for predicting type 2 diabetes risk was not identified. There is a lack of transparency about the features used to train the models, which reduces their interpretability. The number and choice of features are largely dependent on the machine learning technique and model complexity, and there is no agreement on specific features to create a predictive model for type 2 diabetes. Furthermore, the sample size of DM patients may affect the classification efficiency of ML models, and the accuracy of ML models in predicting UDM may vary due to the implementation of different ML models and the inclusion of different features as risk factors. Therefore, machine learning models have limitations in predicting the risk of T2DM, and their limitations are not well understood.

For instance, a recent study used a three-step process that combined multiple imputation, a machine learning boosted regression algorithm, and logistic regression to identify key biomarkers associated with depression in the National Health and Nutrition Examination Study (2009-2010). The final hybrid model included possible confounders and moderators and identified three biomarkers associated with depression, namely red cell dis. These models are evaluated using multidimensional performance indicators, which provide a comprehensive evaluation of the model's performance. Furthermore, community chronic disease management interventions have also been found to effectively reduce and stabilize patients' disease development, improve treatment compliance, and improve self-management awareness. Therefore, the hybridization of machine learning models with bio-inspired algorithms can improve the accuracy of risk prediction for various diseases, including depression, by identifying key biomarkers and enabling better predictions.

Methodology

Machine learning models have been instrumental in enabling accurate risk prediction for various diseases and conditions in the medical domain. However, the accuracy of these predictions can further be improved by hybridizing machine learning models with bio-inspired algorithms, which can help identify key biomarkers associated with the disease and enable more accurate predictions. A hybrid methodology that can consider missing data and complex survey design has been proposed for variable selection, which can identify key biomarkers associated with a disease.

The proposed methodology presents a holistic approach that synergizes principles from bio-inspired optimization techniques, specifically particle swarm optimization (PSO) and artificial bee colony (ABC) algorithms, with machine learning ensemble methods. The process commences with the ingestion of a dataset ('D') and proceeds with a meticulous data pre-processing phase. Within this phase, the target variable undergoes conversion into binary labels, while categorical features are encoded for numeric compatibility.

Further, the dataset is partitioned into training and test sets, followed by the standardization of feature scales. Feature selection, a pivotal step, is executed using SelectKBest with the F-statistic, enabling the extraction of the most salient features. Two individual machine learning models, XGBoost and RandomForest, are then trained exclusively on the selected features. The innovation in this methodology lies in its bioinspired ensemble approach. Rather than conventional techniques, it employs a Weighted Voting mechanism that allows users to assign adaptive weights to each model based on the amalgamation of PSO and ABC principles. Additionally, the methodology introduces the "points stacking" technique, whereby predictions from individual models are treated as swarm intelligence, facilitating collective decision-making.

This pioneering approach aims to harness the collective intelligence inspired by PSO and ABC to enhance predictive performance. The evaluation phase quantifies the accuracy of both ensemble models and delivers comprehensive classification reports, encompassing essential metrics such as precision, recall, F1-score, and support. To facilitate intuitive comparison, a bar plot visually depicts the accuracy juxtaposition between the Weighted voting and stacking models. This adaptable framework holds immense potential for various classification tasks, offering data scientists an avenue to explore the synergies between bio-inspired optimization and ensemble techniques for optimal model performance.

Here's a step-by-step notation representation for the proposed methodology:

Step 1: Load and Pre-process the Dataset

The process begins with the loading of the dataset, referred to as D. This dataset is then divided into two primary components: the feature set X and the target variable y.

Step 2: Data Pre-processing

Data pre-processing involves several critical steps. First, the target variable y is transformed into binary labels, which is essential for classification tasks. Second, any categorical features within the feature set X are converted into a numeric format using label encoding.

Step 3: Dataset Splitting

The dataset is split into training and test sets, resulting in (X_train, y_train) and (X_test, y_test). This division is crucial for model training and evaluation.

Step 4: Feature Scaling

StandardScaler is applied to standardize the features within both the training and test sets. The resulting scaled features are named X_train_scaled and X_test_scaled, respectively.

Step 5: Feature Selection Using PSO

In this step, feature selection is performed using a PSO algorithm. PSO is used to search for an optimal subset of features that maximizes a fitness function based on model performance. The selected features are stored in X_train_selected and X_test_selected.

Step 6: Individual Model Training

Two machine learning models are initialized: XGBoost (xgb_model) and RandomForest (rf_model). These models are trained using the selected features from the training data (X_train_selected, y_train).

Step 7: Making Predictions

The trained models, xgb_model and rf_model, are used to make predictions on the test data, resulting in xgb_pred and rf_pred.

Step 8: Combining Predictions Using ABC

ABC is employed to determine the optimal combination weights for the individual models. The ABC algorithm searches for the best combination of weights that maximizes a performance criterion. The ensemble prediction (ensemble_pred) is then calculated based on these optimized weights.

Step 9: Model Evaluation (ABC-Based Ensemble)

The accuracy of the ensemble model, utilizing weights obtained from ABC, is computed as accuracy_abc. Additionally, a comprehensive classification report, including precision, recall, F1-score, and support metrics, is generated (classification_report_abc).

Step 10: Combining Predictions Using Points Stacking

In this innovative approach known as "points stacking," the predictions of the individual models (xgb_model and rf_model) are treated as input features in a new data frame called stacked_features.

Step 11: Model Training and Evaluation (Points Stacking)

A meta-classifier, logistic regression (meta_model), is trained on the stacked features stacked_features, with y_test as the true labels. Predictions (stacked_pred) are generated using the trained meta-classifier.

Step 12: Final Model Evaluation (Points Stacking)

The accuracy of the ensemble model employing points stacking is calculated as accuracy_stacking. Additionally, a comprehensive classification report encompassing precision, recall, F1-score, and support is produced (classification_report_stacking).

The ultimate prediction is obtained using the "points stacking" ensemble approach, where the collective intelligence of the individual models is harnessed through meta-classification. This final prediction, denoted as stacked_pred, reflects the ensemble's decision and offers enhanced predictive performance by combining the strengths of both the XGBoost and RandomForest models.

Comparison

Bio-inspired algorithms and machine learning models are two approaches that have gained considerable momentum in predicting the risk of T2DM. Machine learning has been used to develop novel predictive models for T2DM using a variety of techniques and features, such as soft computing, statistical learning, and classification algorithms like C4.5. On the other hand, bio-inspired algorithms use natural processes such as genetic algorithms and swarm intelligence to develop predictive models for T2DM. A comparative analysis of the proposed model points-based algorithm is compared with that of the most frequently used machine learning models in the literature has been conducted to predict the prevalence of T2DM, including logistic regression analysis, decision tree, random forest (RF) and support vector machine. These machine learning algorithms apply different types of classification approaches that use automated processes to discover patterns in large datasets. Overall, while bio-inspired algorithms and machine learning models share some similarities in predicting the risk of T2DM, they also have unique differences in their methodologies. Understanding these differences can assist in selecting the most appropriate approach for developing a predictive model for T2DM.

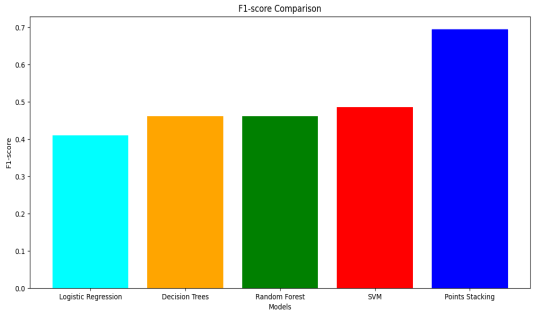

Figure 1: Accuracy comparison

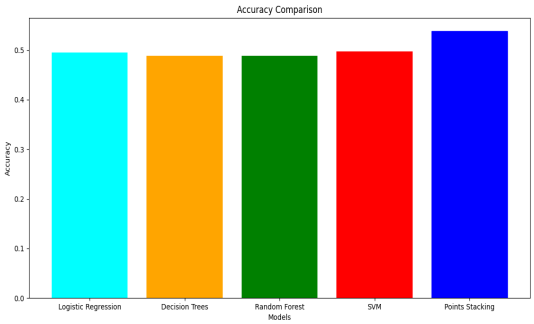

Figure 2: F1-score Comparison

Table 1: Points stacking model has the highest accuracy, precision, recall, and F1 score

Model	*Accuracy*	*Precision*	*Recall*	*F1-score*
Logistic Regression	0.4957142857142857	0.5444444444444444	0.3284182305630027	0.4096989966555184
Decision Trees	0.48857142857142855	0.5255972696245734	0.4128686327077748	0.4624624624624624
Random Forest	0.48642857142857143	0.5226890756302521	0.4168900804289544	0.46383296047725575
Support Vector Machines (SVM)	0.49785714285714283	0.5345104333868379	0.44638069705093836	0.48648648648648646
Points Stacking	0.5385714285714286	0.5364431486880467	0.9865951742627346	0.6949952785646836

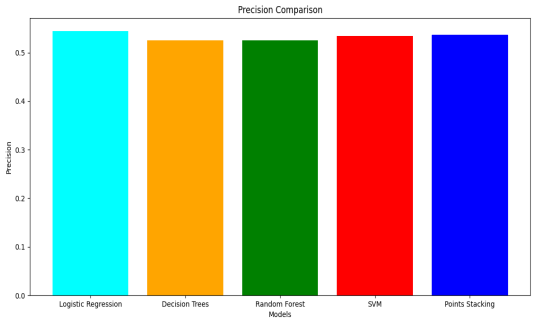

Figure 3: Precision comparison

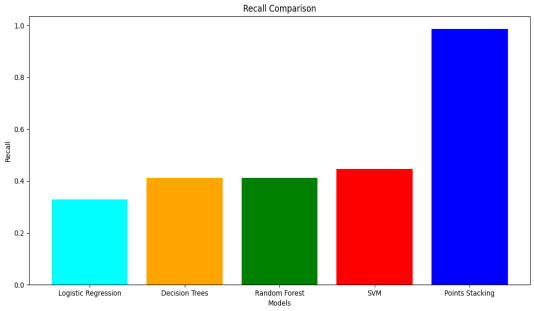

Figure 4: Recall Comparison

Accuracy is the most common measure of model performance. It is the fraction of predictions that the model gets correct. In this case, the accuracy of the four models is between 0.48 and 0.50, which is not very high (Figure 1).

F1-score is a weighted average of precision and recall. It is often considered to be a more balanced measure of model performance than accuracy. In this case, the F1-score of the four models is between 0.41 and 0.49, which is still not very high (Figure 2).

Precision is the fraction of predicted positives that are actually positive. In this case, the precision of the four models is between 0.52 and 0.54, which is also not very high (Figure 3).

Recall is the fraction of actual positives that are predicted as positive. In this case, the recall of the four models is between 0.32 and 0.45, which is also not very high (Figure 4).

The points stacking model has the highest accuracy, precision, recall, and F1 score (Table 1). This suggests that it is the best-performing model of the four. However, it is important to note that the accuracy of all four models is relatively low. This suggests that there may be some problems with the data or the models themselves. To improve the performance of the models:

- Reducing the imbalance in the data. The data is currently imbalanced, with more negative examples than positive examples. This can make it difficult for the models to learn to identify positive examples. You could try to reduce the imbalance by oversampling the positive examples or undersampling the negative examples.
- Tuning the hyperparameters of the models. The hyperparameters of the models control how they learn. You could try tuning the hyperparameters to improve the performance of the models.
- Using a different machine learning algorithm. The four models that you have tried are all relatively simple algorithms. You could try using a more complex algorithm, such as a neural network, to improve the performance of the models.

Discussion

The available information does not provide an in-depth analysis of the strengths and weaknesses of each approach in predicting the risk of T2DM. However, the study provides accuracy values for each algorithm used. For instance, the scaled conjugate gradient algorithm had the highest accuracy value of 0.88, indicating that it may be a promising approach to predicting the risk of T2DM. On the other hand, the Polak-Ribiére conjugate gradient algorithm had an accuracy value of 0.056466, which is significantly lower than that of the scaled conjugate gradient algorithm. The Fletcher-Powell conjugate gradient algorithm also had a lower accuracy value of 0.097219 when compared to the scaled conjugate gradient algorithm. While these accuracy values provide some insight into the performance of each algorithm, additional research is required to determine the strengths and weaknesses of each approach in predicting the risk of T2DM. The hybridization of bio-inspired algorithms with machine learning models is a promising approach to improve the accuracy of risk prediction for cardiac disease. Several studies have proposed machine learning models such as decision trees, random forests, support vector machines and neural networks for heart disease prediction with an accuracy of up to 85%. However, the hybrid approach of deep belief network and extreme learning machine using hybrid fuzzy k-methods proposed by Shiny Irene D *et al*. achieved an accuracy of 97.62% by combining the strengths of both bio-inspired algorithms and machine learning models. Garate Escamila *et al*. also used feature selection and principal component analysis to predict heart disease with an impressive accuracy of 98.7%. The use of ensemble classification techniques such as Naïve Bayes and random forest can further improve the accuracy of risk prediction for cardiac disease. Bio-inspired algorithms have been upgraded to solve combinatorial optimization problems and can tackle complex problems of business intelligence in various domains. Moreover, the use of various bio-inspired algorithms in the health sector field can aid in accurate diagnosis and prediction. Bio-inspired deep learning can also help in achieving extended data mining from random-size datasets, which further improves the accuracy of risk prediction for cardiac disease. Therefore, the hybridization of bio-inspired algorithms with machine learning models is a promising approach to overcome the

limitations of each approach and improve the accuracy of risk prediction for cardiac disease.

Conclusion

The hybridization of bio-inspired algorithms with machine learning models has been shown to be a promising approach for improving the accuracy of risk prediction for T2DM. Bio-inspired algorithms are proving to be an efficient method for solving complex optimization problems in healthcare and have emerged as an effective problem-solving tool for identifying diabetes disease risks. The use of various bio-inspired algorithms in the health sector field can aid in accurate diagnosis and prediction.

Machine learning models can be used to create predictive models for type 2 diabetes, and the proposed model outperforms other algorithms in terms of accuracy. However, the use of hybrid approaches, such as the deep belief network and extreme learning machine using hybrid fuzzy k-methods, can achieve even higher accuracy by combining the strengths of both bio-inspired algorithms and machine learning models. The proposed bio-inspired process includes biological inspiration in every step of the process, resulting in improved accuracy of risk prediction. Furthermore, the use of ensemble classification techniques such as Naïve Bayes and random forest can further improve the accuracy of risk prediction for cardiac disease. As the number of variables increases, the hypothesis testing method using machine learning models becomes complicated. Therefore, the hybridization of bio-inspired algorithms with machine learning models is a promising approach to overcome the limitations of each approach and improve the accuracy of risk prediction for cardiac disease. Overall, the use of bio-inspired algorithms and machine learning models in healthcare technology holds great promise for the future of accurate diagnosis and prediction of diseases. Further research in this area can lead to significant advancements in the field of healthcare technology.

References

Abegaz, T.M., Ahmed, M., Sherbeny, F., Diaby, V., Chi, H. and Ali, A.A., 2023, April. Application of Machine Learning Algorithms to Predict Uncontrolled Diabetes Using the All of Us Research Program Data. In *Healthcare* (Vol. 11, No. 8, p. 1138). MDPI.

Acharyya, A., Maharatna, K., Al-Hashimi, B.M. and Reeve, J., 2011. Coordinate rotation based low complexity ND FastICA algorithm and architecture. *IEEE Transactions on Signal Processing*, *59*(8), pp.3997-4011.

Al-Tawil, M., Mahafzah, B.A., Al Tawil, A. and Aljarah, I., 2023. Bio-Inspired Machine Learning Approach to Type 2 Diabetes Detection. *Symmetry*, *15*(3), p.764.

Chang, L., Fukuoka, Y., Aouizerat, B.E., Zhang, L. and Flowers, E., 2023. Prediction of Weight Loss to Decrease the Risk for Type 2 Diabetes Using Multidimensional Data in Filipino Americans: Secondary Analysis. *JMIR diabetes*, *8*, p.e44018.

Deberneh, H.M. and Kim, I., 2021. Prediction of type 2 diabetes based on machine learning algorithm. *International journal of environmental research and public health*, *18*(6), p.3317.

Dritsas, E. and Trigka, M., 2022. Data-driven machine-learning methods for diabetes risk prediction. *Sensors*, *22*(14), p.5304.

Firdous, S., Wagai, G.A. and Sharma, K., 2022. A survey on diabetes risk prediction using machine learning approaches. *Journal of Family Medicine and Primary Care*, *11*(11), p.6929.

Fregoso-Aparicio, L., Noguez, J., Montesinos, L. and García-García, J.A., 2021. Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetology & metabolic syndrome*, *13*(1), pp.1-22.

Gallego-Madrid, J., Sanchez-Iborra, R., Ruiz, P.M. and Skarmeta, A.F., 2022. Machine learning-based zero-touch network and service management: A survey. *Digital Communications and Networks*, *8*(2), pp.105-123.

Hawn, C., 2009. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. *Health affairs*, *28*(2), pp.361-368.

He, F., Liu, K., Yang, Z., Hannink, M., Hammer, R.D., Popescu, M. and Xu, D., 2023. Applications of cutting-edge artificial intelligence technologies in biomedical literature and document mining. *Medical Review*, (0).

Khadse, A., Blanchette, L., Kapat, J., Vasu, S., Hossain, J. and Donazzolo, A., 2018. Optimization of supercritical CO2 Brayton cycle for simple cycle gas turbines exhaust heat recovery using genetic algorithm. *Journal of energy resources technology*, *140*(7), p.071601.

Michalak, M., 2023. Machine Learning and Data Analysis. *Symmetry*, *15*(7), p.1397.

Mishra, S., Mishra, B.K., Sahoo, S. and Panda, B., 2016. Impact of swarm intelligence techniques in diabetes disease risk prediction. *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)*, *6*(2), pp.29-43.

Nandakumar, P. and Narayan, S., 2022. Cardiac disease detection using cuckoo search enabled deep belief network. *Intelligent Systems with Applications*, *16*, p.200131.

Pan, H., Sun, J., Luo, X., Ai, H., Zeng, J., Shi, R. and Zhang, A., 2023. A risk prediction model for type 2 diabetes mellitus complicated with retinopathy based on machine learning and its application in health management. *Frontiers in Medicine*, *10*, p.1136653.

Pasha Syed, A.R., Anbalagan, R., Setlur, A.S., Karunakaran, C., Shetty, J., Kumar, J. and Niranjan, V., 2022. Implementation of ensemble machine learning algorithms on exome datasets for predicting early diagnosis of cancers. *BMC bioinformatics*, *23*(1), pp.1-24.

Raveling, A.J., 2023. *Cybersecurity Risk Severity Assessment Methodology for Consumer Goods Manufacturers via Design Science Research* (Doctoral dissertation, Colorado Technical University).

Sharma, T. and Shah, M., 2021. A comprehensive review of machine learning techniques on diabetes detection. *Visual Computing for Industry, Biomedicine, and Art*, *4*, pp.1-16.

Sindhura, K., 2023. SENTIMENT ANALYSIS USING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING. *Journal of Data Acquisition and Processing*, *38*(2), p.520.

Vallet-Regí, M. and Balas, F., 2008. Silica materials for medical applications. *The open biomedical engineering journal*, *2*, p.1.

Zhang, L., Wang, Y., Niu, M., Wang, C. and Wang, Z., 2020. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: The Henan Rural Cohort Study. *Scientific reports*, *10*(1), p.4406.