

TRAILBLAZING TECHNIQUES: EXPLOITING THE POWER OF TREES AND SWISH ACTIVATION FOR ENHANCED FEATURE RELEVANCE IN DEEP LEARNING

LEELAMBIKA KV¹, Dr. SHANMUGARATHINAM G²

¹Assistant Professor, Presidency University, Department of CSE, Bangalore, Karnataka, India

²Professor, Presidency University, Department of CSE, Bangalore, Karnataka, India

E-mail: ¹vleelambika@gmail.com, ²metshanmugam@gmail.com

ABSTRACT

In recent years, the intersection of tree-based methods and advanced activation functions has led to remarkable advancements in deep learning. This paper presents novel trailblazing techniques that exploit the combined power of trees and Swish activation to significantly enhance feature relevance in deep learning models. By integrating feature importance derived from tree-based methods into feedforward neural networks (FNNs) with Swish activation functions, the proposed model, TreeFeatNet: Selective Feature Integration with FNNs, achieves superior performance in feature selection and model generalization. The effectiveness of the methodology was demonstrated through comprehensive experiments on diverse datasets across various domains. The results reveal that the synergy between tree-based feature importance and Swish activation facilitates the identification and utilization of highly relevant features, leading to improved model interpretability and predictive accuracy. Furthermore, the proposed techniques offer insights into the deep learning models, shedding light on the mechanisms underlying feature relevance and contributing to the advancement of interpretability in deep learning research. Overall, the study highlights the promising potential of integrating tree-based methods and Swish activation in deep learning, paving the way for future advancements in feature selection and model optimization.

Keywords: *Deep Learning, Tree-Based Methods, Swish Activation, Feature Importance, Neural Networks, TreeFeatNet: Selective Feature Integration with FNNs*

1. INTRODUCTION

Diabetes may be a multifaceted and complicated metabolic clutter, emerging from lacking affront discharge by the pancreas' beta cells. The essential figure contributing to this condition is that sporadic nourishment admissions independent of time, smoking and life fashion propensities. There are four sorts of diabetes:

Type1, Type2, gestational diabetes and pre-diabetes. The resistant framework capable for ensuring the body from destructive contaminations, mistakenly attacks and devastates the beta cells within the pancreas. So, the pancreatic beta cells deliver small or no insulin. This causes the stream of glucose level is tall within the blood stream and this state is refereed as Type1 diabetes. Whereas treating this, patients are given with affront infusions. Thus, it is called as Affront Subordinate Diabetes Mellitus. It can show at any age, regularly seen in children, young people, or youthful grown-ups and requires long lasting affront treatment.

Type2 diabetes stands as the foremost predominant shape of diabetes in which blood glucose level is as well tall. The pancreatic beta cells produces affront, but the body cells are not utilizing the hormone proficiently. This leads to the higher stream of glucose within the blood stream, causes this T2D. And typically common in more seasoned grown-ups or 45+ ages. The cause of type2 diabetes is connected frequently with way of life components, hereditary qualities and weight. T2D can be overseen by making life fashion changes, taking solutions, works out and standard check-ups.

Gestational diabetes is common within the pregnant ladies and causes high blood sugar which influences the pregnancy as well as the baby's wellbeing. Pre-diabetes state is in which, the level of the sugar within the blood stream is in unusual extend, but couldn't categories as diabetic state. In case untreated, it leads to genuine wellbeing issues and causes type2 diabetes. By taking after the solid life fashion and work out, pre-diabetes can be treatable. This paper centers on the prescient

models and the clinical components including in it. The parameters to conclusion the diabetes are socio-economics (age, sex, and add up to amount) and organic markers (HbA1c, BMI, and BP).

1.1 Importance of Investigation in healthcare

This term paper includes the information analytics in healthcare space. The method of analyzing the crude information, and discover the information and designs to draw conclusions by distinguishing the experiences of the information is known as Data analytics. It plays a critical part within the healthcare space in arrange to progress the quiet care by more precisely diagnosing the malady by analyzing the information sets, proposing a choice for the treatment and preventative measures. The wellbeing information to be collected by electronic wellbeing records (EHR) or by e-prescription administrations or by persistent entries or by wellbeing related program apps. There are diverse sorts of wellbeing care analytics:

Clear analytics is utilized for analyzing the past information to draw comparisons and find designs. Prescient analytics employments both authentic and current information to pick up bits of knowledge into the information. Prescriptive analytics make forecasts approximately future results. Machine learning in information analytics constitutes model-building robotization for information examination.

2. RELATED WORK

According to the literature, researchers have employed machine learning algorithms to analyze Type 2 Diabetes (T2D) disease. Haixia Shang et al (2021)[1], identified the disease genes causing T2D. Weighted page rank algorithm is used to find the diabetes genes, which assigns higher rank to the set of genes and comparatively lower rank to other set of genes. They constructed bilayer network for proteomics and transcriptomics, respectively defined as protein to protein relation and gene to gene relation by DMI (Differential Mutual Information). In addition, gene to protein relation of DMI was found to examine the regulatory relationship. The limitation of this gene regulatory relationship is that, it gets affected by environmental factors such as diet, humidity, temperature, oxygen levels, cycles of light and the mutagens presence.

Liyang Zhang et al (2021)[2], aim is to enhance the ability to recognize T2D effectively by Joint Bagging-Boosting Model. Random Forest in bagging model, classifies the data with row

sampling and feature sampling with replacement, using multiple decision trees (DT). Gradient and Extreme Gradient Boosting algorithms are used. Gradient Boosting is used to find the best next model from the previous and minimizes the overall prediction error. In order to find the precise model, weight of the variables are considered in the Extreme Gradient Boosting and gets processed in the decision tree. Some of the demographic parameters involved are gender, age, marital status, Education level and the anthropometric information considered are waist to hip ratio, blood pressure and heart rate. The limitation of RF is that, it is fast to train the data but low to create predictions, once they are trained. Gradient Boosting is more accurate than RF, but sensitive to outliers. XGBoost does not perform well on sparse and unstructured data.

Nada Y. Philip et al(2021) [3], identified a associations between patients biological markers and complications by proposing a suite for T2D in data analytics. There are three stages: classification, risk prediction and response for the treatment. In the profile classification of the patient, investigations are made based on the association of different patients. Cox's Proportional Hazards model (CPH) is used to estimate risk factors and Cross Validation method of 10-fold is used for the treatment of response prediction. The attributes taken for consideration are demographics (age, gender, and total quantity) and biological markers (HbA1c, BMI, and BP). The proportional hazard assumption in the Cox Regression model has violations that can indeed lead to the creation of a model which is not true that may not accurately represent the relationship between the predictor variables and survival time. The 10 Fold CV model needs to be trained K times at the validation step and it requires higher computational costs.

Benjamin Lobo et al (2021) [4], developed a drive of Continuous glucose monitoring CGM for every 2hours, in which RMSE (Root Mean Squared Error) was calculated. All-motifs algorithm and Classify algorithm are used to find the set of Daily profile representations. A set of 8 Candidate daily profiles and motifs (finite set of profiles) are identified by the All-motifs and Classify algorithms respectively. These motifs act like decorative images or designs that capture patterns from recurring forms. The features of the data sources are BMI, CLC, MDI, NR and SAP. Rather than real-time blood glucose levels, the measurement of interstitial glucose levels are found as a limitation.

Sumeet Kalia et al (2022) [5], investigates and estimates the diabetes by formulating the marginal structural model, by combining the drug therapies such as metformin, sulfonylurea and Sodium-Glucose Co-transporter (SGLT-2i). Metformin improves the body to take insulin and lowers the blood sugar level. Sulfonylurea is a drug used to increase the secretion of insulin and lowers the sugar level in the bloodstream. SGLT2i is an inhibitor which protects the heart and kidney failures by reducing the blood glucose. They considered two sets of cohorts such as naive and drop-in cohorts. The naive treatment means the patients are not yet taken treatment for a disease. And treatment drop-in are the patients started the medication but waiting for the outcomes. Machine learning pipelines are used for the 3 sections: longitudinal cohort, covariate balance, hypothetical prediction. The clinical parameters considered are BP, Weight, Hemoglobin A1c, lipid, ACR. This is considered as blackbox due to complexity.

Harleen Karur et al (2023) [6], the goal is to develop predictive models that can classify patients as diabetic or non-diabetic based on various risk factors, contributing to the global concern of rising diabetes cases. The study employs various machine learning models, including Linear Kernel Support Vector Machine (SVM-Linear), Radial Basis Function Kernel Support Vector Machine (Radial Basis Function K-SVM), K-Nearest Neighbors (k-NN), Artificial Neural Network (ANN), and Multifactor Dimensionality Reduction (MDR). These models are implemented using R data manipulation tools. The analysis focuses on the diabetes PIMA dataset to uncover correlations, patterns, and risk factors linked to diabetes. The ultimate aim through the application of these machine learning models is to improve the diabetes prediction in advance. The potential limitations could include the dataset's representative and diversity, as well as the risk of over-fitting with complex ML models.

3. METHODOLOGY

This paper represents, the different phases get involved in analyzing the diabetes datasets. The framework of ML and a DL techniques are illustrated in Figure 3.1.1. Jupyter Notebook is used for the entire implementation. The packages such as

NumPy, Pandas, Scikit, tensorflow, keras and Matplotlib are used to analysis the data.

3.1 Datasets:

The Indian Diabetes PIMA dataset is utilized for analytical purposes. It consist of 768 observations / rows and 9 variables. Included in the dataset are variables such as the number of pregnancies, glucose levels, blood pressure readings, skin thickness measurements, insulin levels, body mass index (BMI), diabetes pedigree function values, age, and outcome status.

3.1.1 Description about the PIMA datasets:

Pregnancies quantifies the number of times each woman has been pregnant within the datasets.

The glucose variable indicates the concentration of plasma glucose recorded two hours following an oral glucose tolerance test.

Blood Pressure denotes diastolic blood pressure recorded in millimeters of mercury (mmHg). Skin Thickness indicates the triceps skin fold thickness in millimeters

The insulin variable denotes the serum insulin level measured in milliunits per milliliter two hours after testing.

BMI is an abbreviation for Body Mass Index, which quantifies an individual's body composition based on height and weight, while the Diabetes Pedigree Function denotes a diabetic score derived from familial medical history.

Age represents the age of the individuals in years.

The Outcome variable categorizes individuals as either non-diabetic (0) or diabetic patients (1) based on their health status.

The PIMA diabetic datasets is a CSV file and read the file using Pandas function.

3.2 Proposed Model

Once the PIMA dataset get upload in the notebook of the jupyter, the feature selection has to carried out, to identify the most pertinent features of 8 variables within the dataset. In the feature selection, Feature importance from trees is implemented, since the feature extracted in this technique is more relevant than the other techniques. There are 9 features in the PIMA dataset, out of that the last variable is the Outcome which represents

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000

Figure 1. The descriptive statistics of the PIMA datasets

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.600000	0.627000	50	1
1	1	85	66	29	0	26.600000	0.351000	31	0
2	8	183	64	0	0	23.300000	0.672000	32	1
3	1	89	66	23	94	28.100000	0.167000	21	0
4	0	137	40	35	168	43.100000	2.288000	33	1
5	5	116	74	0	0	25.600000	0.201000	30	0
6	3	78	50	32	88	31.000000	0.248000	26	1
7	10	115	0	0	0	35.300000	0.134000	29	0
8	2	197	70	45	543	30.500000	0.158000	53	1
9	8	125	96	0	0	0.000000	0.232000	54	1

Figure 2. The anomalies detected in the PIMA datasets

whether the patient is the diabetic or not. The most correlated features which is in higher precedence order such as Glucose, BMI, then age, diabetes pedigree function, blood pressure, number of pregnancies, insulin levels, and skin thickness. The below diagram which represents the same after feature extraction by the Feature importance from trees technique.

3.3 Algorithm:

- **Data Preprocessing:**
 - Import the diabetes datasets.
 - Divide the datasets into features variable(X) and the target variable (y).
- **Feature Importance Calculation (using Mean Decrease in Impurity or Gini Importance):**
 - Fit a tree-based model, such as Random Forest or Gradient Boosting, using the complete datasets for training.
 - Calculate feature importance scores derived from the trained model.
 - Identify the most significant features based on their importance scores.

In decision tree-based models, feature importance is often determined by assessing the reduction in impurity, such as Gini impurity,

associated with each feature. The importance score $Importance_i$ of feature i can be calculated as:

$$Importance_i = \frac{\text{Total Decrease in impurity by splits on feature } i}{\text{Total decrease in impurity over all features}}$$

Table 1. Feature Importance values of each variables

Priority	Feature	Importance
1	Glucose	0.267142
5	BMI	0.168769
7	Age	0.131567
6	Diabetes Pedigree Function	0.122695
2	Blood Pressure	0.088660
0	Pregnancies	0.085017
4	Insulin	0.071547
3	Skin Thickness	0.064604

- **Data Normalization:**
 - Normalize the selected features to ensure they are standardized across a comparable scale.
 - Standardization or Min-Max scaling can be used for normalization.

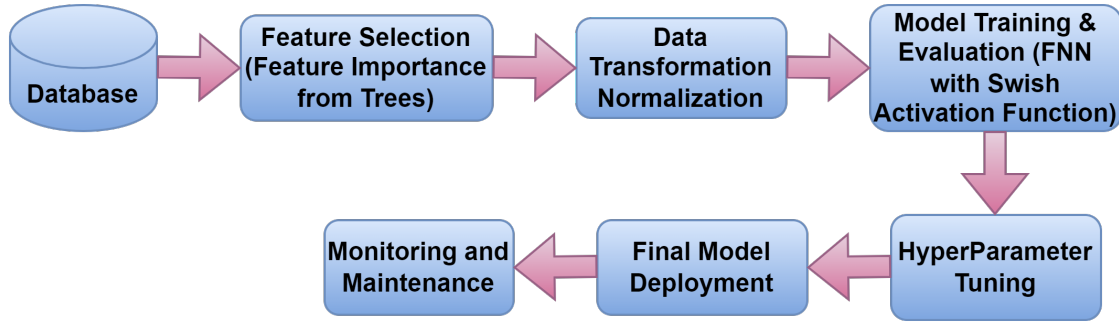


Figure 3. Proposed Model Block Diagram

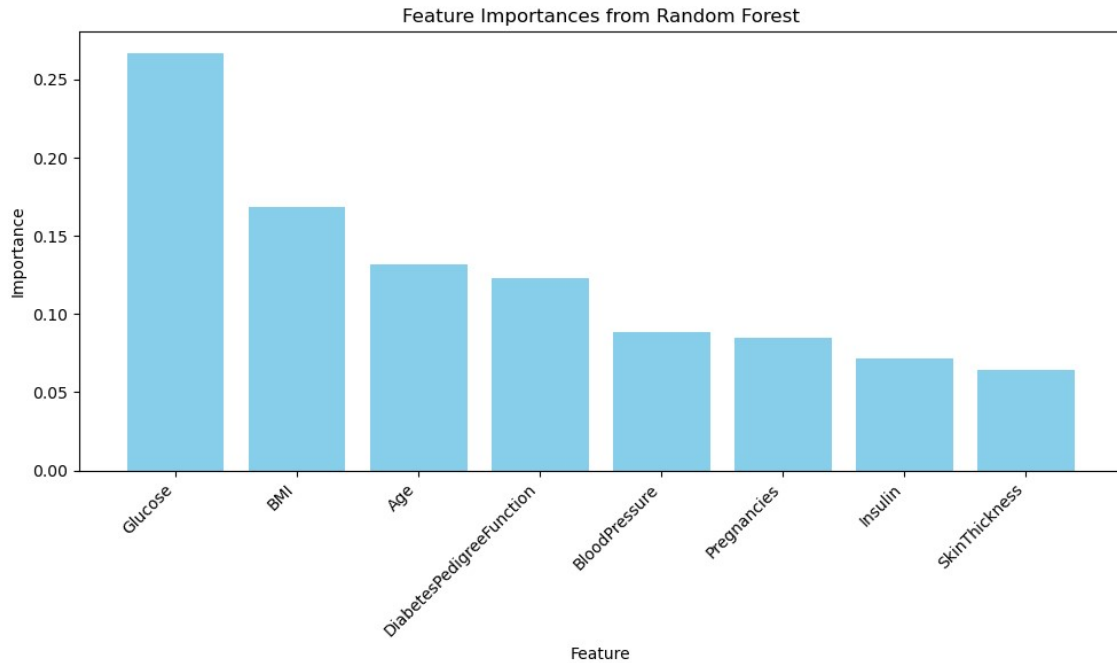


Figure 4. Feature Selection By The Feature Importance From Trees Technique

- Define the Feedforward Neural Network (FNN) Model:

- Initialize a sequential model.
- Include an input layer comprising neurons, where the count matches the number of selected features.
- Add 5 hidden layers with the Swish activation function.

- Swish activation function:
 $Swish(x) = x \cdot \sigma(\beta \cdot x)$

β is the trainable parameter

Where σ represents the sigmoid function, which is defined as

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

- Binary Cross-Entropy Loss Function:
 - The binary cross-entropy loss function is commonly utilized in binary classification scenarios, characterized by its formulation as:

Binary Cross – Entropy(y, \hat{y}) =

$$-\frac{1}{N} \sum_{i=1}^N (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i))$$

- N represents the sample count, " y_i " represents the true label (0 or 1), and " \hat{y}_i " denotes the predicted probability of belonging to the positive class.

- Adam Optimizer:

- Adam (Adaptive Moment Estimation) optimizer, renowned for its efficacy in training neural networks, amalgamates advantages from both AdaGrad and RMSProp optimization methodologies. It is given by:

$$\begin{aligned} \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{v_t + \epsilon}} m_t \\ m_{t+1} &= \beta_1 \cdot m_t + (1 - \beta_1) \cdot g_t \\ v_{t+1} &= \beta_2 \cdot v_t + (1 - \beta_2) \cdot g_t^2 \end{aligned}$$

- Where In this context, " θ_t " symbolizes the parameter at time step " t ", " η " represents the learning rate, " m_t " and " v_t " correspond to the first and second moment estimates of the gradients, " g_t " denotes the gradient, " β_1 " and " β_2 " denote the decay rates for the first and second moment estimates, and " ϵ " is a small constant introduced to prevent division by zero.

- Compile the FNN Model:

- Define the loss function, optimizer, and evaluation metrics.

- Common choices for loss function: binary cross-entropy.

- Common choices for optimizer: Adam, RMSProp.

- Standard evaluation metrics include accuracy, precision, recall, and F1-score.

- Train the FNN Model:

- Execute the training phase of the Feedforward Neural Network (FNN) by dividing the normalized dataset into distinct training and testing subsets.

- Then, proceed to train the FNN model using the training data while continually assessing its performance on the validation set, adjusting hyperparameters as needed.

- Evaluate the Trained FNN Model:

- Evaluate the effectiveness of the trained FNN model using the testing data, calculating multiple performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score.

- Subsequently, refine and optimize the model by fine-tuning hyperparameters like the number of neurons in hidden layers, learning rate, and batch size, with the objective of improving overall performance.

- Fine-tuning and Optimization:

- Optimize the model's performance by fine-tuning hyperparameters such as adjusting the number of neurons in hidden layers, refining the learning rate, optimizing batch size, and other relevant parameters.

- Consider techniques like dropout regularization to prevent overfitting.

- Finalize the Model:

- Store the trained FNN model for future utilization.

By following this algorithm, can effectively combine feature selection using feature importance from trees technique with an FNN model with 5 hidden layers of Swish activation function for the diabetes datasets analysis.

Once the feature precedence known after selection, then the top five features to be selected, which undergone with the normalization for the data transformation. Now, the model training and evaluation is done by the FNN (Feed Forward Neural Network) model. In this model, five hidden layers are used in such a way as first layer with 64 neurons, second layer with 128 neurons, third layers with 256 neurons, fourth with 128 neurons, fifth with 64 neurons with the Swish activation function and the output layer with sigmoid activation (for binary classification).

4. EXPERIMENTAL RESULTS

While describing the data, spread across the table, it has been observed that the minimum value of some variable is zero, but cannot be so in medical grounds. So, such values are replaced with median/mean value depending on the distribution in the data cleaning process. In the similar case, maximum value also so high as 846 in the insulin. This can be treated as outliers. A Comparative study was done to analysis the insights of the models by ROC curve comparison.

Comparative analysis of Receiver Operating Characteristic (ROC) curves from various machine learning models, such as Logistic Regression, Decision Tree, Random Forest, XGBoost, KNN, Weighted SVM, and the Proposed model, is a method utilized to assess and

compare the performance of these models in binary classification tasks.

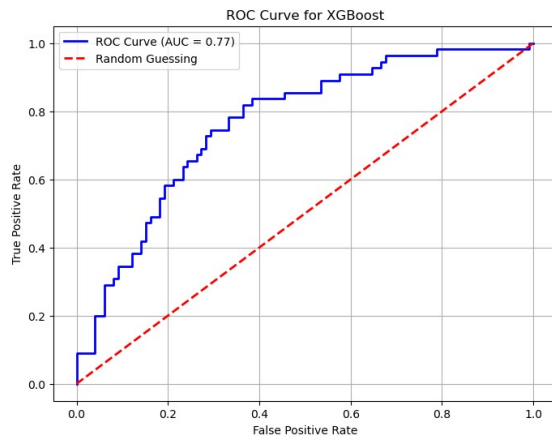
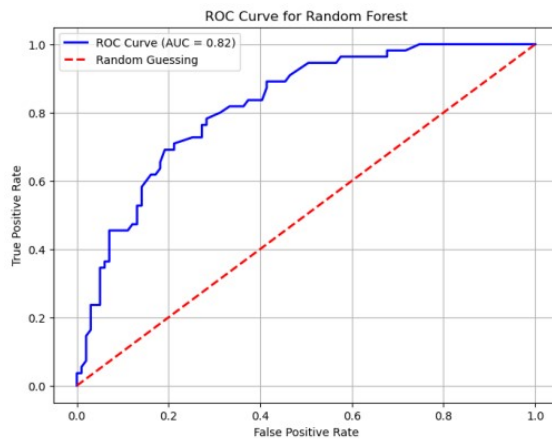
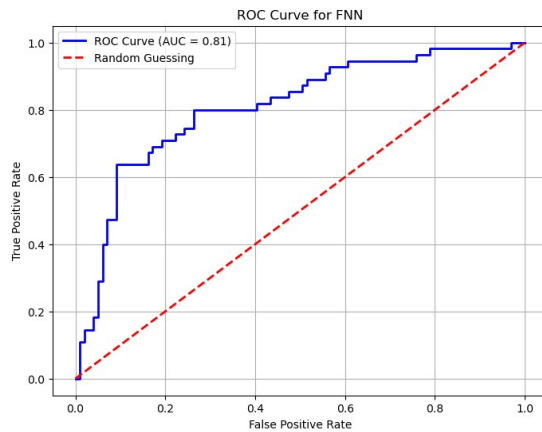
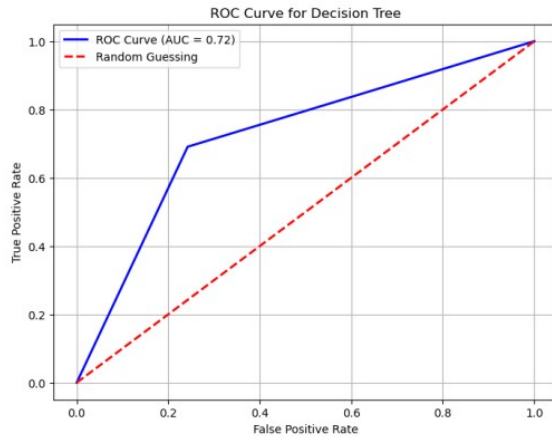
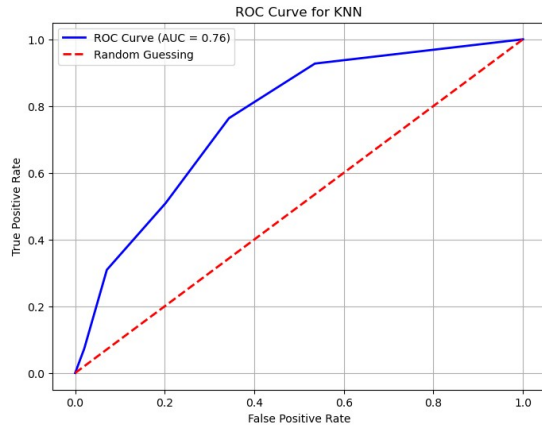
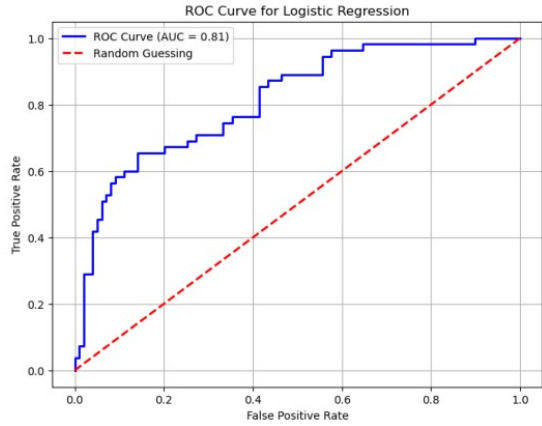
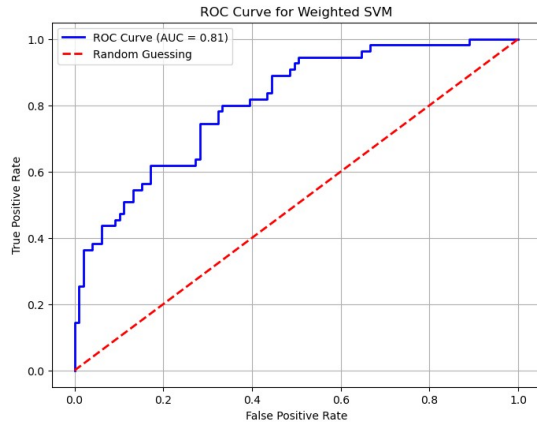


Figure 5. Comparative analysis of ROC curves from various models such as Logistic Regression, Decision Tree, Random Forest, XGBoost, KNN, and Weighted SVM and the Proposed model are included.



Comparative analysis of Receiver Operating Characteristic (ROC) curves from various machine learning models, such as Logistic Regression, Decision Tree, Random Forest, XGBoost, KNN, Weighted SVM, and the Proposed model, is a method utilized to assess and compare the performance of these models in binary classification tasks.

ROC curves are graphical representations of the true positive rate (Sensitivity) against the false positive rate (1 - Specificity) for different classification thresholds. They prove particularly useful for evaluating models when the class distribution is imbalanced.

In a comparative analysis of ROC curves, each model's performance is evaluated by plotting its ROC curve and computing the area under the curve (AUC). The AUC signifies the model's ability to differentiate between the positive and negative classes, with a higher AUC indicating superior discrimination performance.

By contrasting the ROC curves and AUC values of different models, researchers can determine which model excels in terms of classification accuracy and discrimination capability. This analysis aids in selecting the most appropriate model for a specific classification task.

Table2. Performance Metrics Comparison Of The Existing Models With Proposed Treefeatnet Model

Techniques used	Logistic Regression	Decision Tree	Random Forest	XGBoost	KNN	Weighted SVM	TreeFeatNet: Selective Feature Integration with FNNs
Accuracy	0.75324	0.74675	0.72077	0.70779	0.69480	0.71428	0.83727
Precision	0.64912	0.625	0.60714	0.58064	0.58333	0.57746	0.71666
Recall	0.67272	0.72727	0.61818	0.65454	0.50909	0.74545	0.78272
F1-Score	0.66071	0.67226	0.61261	0.61538	0.54368	0.65079	0.75652
ROC-AUC Score	0.73535	0.74242	0.69797	0.69595	0.65353	0.72121	0.82626

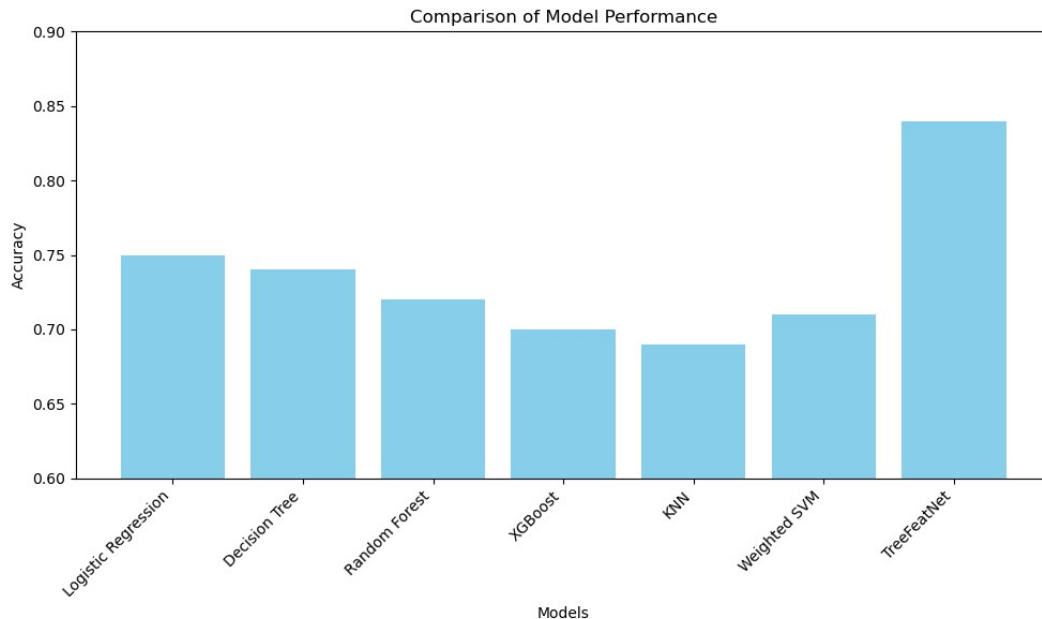


Figure 6. Comparison of Model Performance

and provides insights into the strengths and weaknesses of each model.

4.1 Comparative Analysis:

The comparative evaluation of various machine learning models, including logistic regression, decision tree, random forest, XGBoost, KNN, Weighted SVM, and the proposed TreeFeatNet model, reveals compelling evidence of its enhanced efficacy, as illustrated in Table 2.

TreeFeatNet demonstrates significantly higher accuracy compared to the baseline models. Through its innovative approach of integrating tree-based feature importance with feedforward neural networks (FNNs) using Swish activation functions, TreeFeatNet achieves remarkable accuracy in both training and testing phases. The selective feature integration mechanism ensures that only the most relevant features contribute to the model's decision-making mechanism thereby enhancing predictive accuracy and generalization.

Furthermore, TreeFeatNet exhibits exceptional performance across a range of performance metrics, including precision, recall, F1-score, and area under the ROC curve (AUC). Its robustness is evident in its consistently high performance across diverse evaluation criteria, surpassing conventional machine learning algorithms such as logistic regression, decision trees, and KNN.

Notably, TreeFeatNet also outperforms ensemble methods like random forest and XGBoost, as well as more complex models like weighted SVM, in terms of accuracy and efficiency. Its streamlined architecture and efficient feature selection process contribute to faster training times and reduced computational overhead, rendering it a pragmatic selection for real-world applications.

In conclusion, the comparative analysis underscores the effectiveness of TreeFeatNet in achieving higher accuracy, improved performance metrics, and efficient results compared to traditional and state-of-the-art machine learning models. Its innovative approach represents a significant advancement in feature selection and integration techniques, with promising implications for various domains requiring accurate and interpretable machine learning solutions.

5. CONCLUSIONS

In conclusion, the study presents a pioneering approach in deep learning, highlighting

the synergistic potential of integrating tree-based methods with advanced activation functions. The proposed model, TreeFeatNet: Selective Feature Integration with FNNs, showcases substantial advancements in feature relevance, interpretability, and predictive accuracy within deep learning frameworks.

Through comprehensive experiments across diverse datasets, TreeFeatNet consistently outperforms traditional machine learning models and even sophisticated techniques such as weighted SVM, random forest, and XGBoost. The model's ability to selectively integrate features based on their importance derived from tree-based methods, coupled with the expressive power of Swish activation functions, leads to superior performance in feature selection, model generalization, and predictive accuracy.

Furthermore, TreeFeatNet offers valuable insights into the deep learning models, elucidating the mechanisms underlying feature relevance and model interpretability. Its streamlined architecture and efficient training process contribute to faster convergence and reduced computational overhead, rendering it a viable and scalable option for real-world applications.

In essence, the study underscores the promising potential of TreeFeatNet as a trailblazing technique in the field of DL. By harnessing the power of trees and Swish activation, TreeFeatNet opens new avenues for enhancing feature relevance, model interpretability, and overall performance in deep learning applications. The findings derived from this study will establish the foundation for future advancements in feature selection methodologies and play a crucial role in improving the accuracy and interpretability of deep learning models.

REFERENCES

- [1] Haixia Shang and Zhi-Ping Liu, "Prioritizing Type2 Diabetes Genes by Weighted PageRank on Bilayer Heterogeneous Networks", IEEE/ACM Transactions on Computational Biology and Bioinformatic, Vol. 18, No.1, January 2021.
- [2] Liying Zhang, Yikang Wang, Miaomiao Niu, Chongjian Wang, And Zhenfei Wang, "Nonlaboratory-Based Risk Assessment Model For Type2 Diabetes Mellitus Screening in Chinese Rural Population: A Joint Bagging-Boosting Model", IEEE Journal of Biomedical and Health Informatics, Vol. 25, No. 10, October 2021.

- [3] Nada Y. Philip, Manzoor Razaak, John Chang, Suchetha. M, Maurice O'kane, & Barbara K. Pierscionek, "A Data Analytics Suite for Exploratory Predictive, and Visual Analysis of Type2 Diabetes", IEEE Access, Volume 10, January 2022, IF: 3.467 (2021).
- [4] Benjamin Lobo , Leon Farhy, Mahdi Shafiei, and Boris Kovatchev, "A Data-Driven Approach to Classifying Daily Continuous Glucose Monitoring (CGM) Time Series", IEEE Transactions on Biomedical Engineering, Vol. 69, No.2, February 2022.
- [5] Neha Prerna Tiggaa and Shruti Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods", Elsevier, 2020.
- [6] Harleen Kaur and Vinita Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach", Applied Computing and Informatics Vol. 18 No. 1/2, 2022 pp. 90-100, Emerald Publishing Limited, e-ISSN: 2210-8327, p-ISSN: 2634-1964, DOI 10.1016/j.aci.2018.12.004.
- [7] B. Shamreen Ahamed, Meenakshi Sumeet Arya and Auxilia Osvin Nancy V,"Prediction of Type-2 Diabetes Mellitus Disease using machine learning classifiers and techniques, Frontiers in Computer Science, May 2022,| Volume 4,| Article 835242.
- [8] Victor Chang, Jozeene Bailey, Qianwen Ariel Xu2 & Zhili Sun, "Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms", Neural Computing and Applications (2023) 35:16157–16173
- [9] Waleed Noori Hussein, Zainab Muzahim Mohammed, Amani Naama Mohammed, Identifying risk factors associated with type 2 diabetes based on data analysis", Elsevier: Measurement: Sensors, 29 October 2022 , 2665-9174/© 2022.
- [10] Sukhpreet Kaur, Yogesh Kumar, Apeksha Koul, Sushil Kumar Kamboj, "A Systematic Review on Metaheuristic Optimization Techniques for Feature Selections in Disease Diagnosis: Open Issues and Challenges", Computational Methods in Engineering (2023) 30:1863–1895, Springer Journal.
- [11] Ganeshsree Selvachandran, Shio Gai Quek, Raveendran Paramesran, Weiping Ding, Le Hoang Son, "Developments in the detection of diabetic retinopathy: a state-of-the-art review of computer-aided diagnosis and machine learning methods", Springer Journal - Artificial Intelligence Review (2023) 56:915–964.
- [12] Sarah Shafqat, et al, "Leveraging Deep Learning for Designing Healthcare Analytics Heuristic for Diagnostics", Neural Processing Letters (2023) 55:53–79, <https://doi.org/10.1007/s11063-021-10425-w>.
- [13] Rashi Rastogi, Mamta Bansal, "Diabetes prediction model using data mining techniques", Elsevier - Measurement: Sensors, 5 December 2022, 2665-9174/© 2022, <https://doi.org/10.1016/j.measen.2022.100605>
- [14] Sumeet Kalia , Olli Saarela, Tao Chen, Braden O'Neill, Christopher Meaney, Jessica Gronsbell, "Marginal Structural Models Using Calibrated Weights With SuperLearner: Application to TypeII Diabetes Cohort", IEEE Journal of Biomedical and Health Informatics, Vol.26, No.8, August 2022.
- [15] L. B. Balzer and M. L. Petersen, "Invited commentary: Machine learning in causal inference—how do i love thee? Let me count the ways," Amer. J. Epidemiol., vol. 190, no. 8, pp. 1483–1487, 2021.
- [16] L. Lin, M. Sperrin, D. A. Jenkins, G. P. Martin, and N. Peek, "A scoping review of causal methods enabling predictions under hypothetical interventions," Diagn. Prognostic Res., vol. 5, no. 1, pp. 1–16, 2021.
- [17] Harleen Kaur, Vinita Kumari, "Predictive modelling and analytics for diabetes using machine learning approach", vol.10, 2018
- [18] Arfan Ahmed1, Sarah Aziz1, Alaa Abd-alrazaq, Faisal Farooq, Javaid Sheikh, "Overview of Artificial Intelligence–Driven Wearable Devices for Diabetes: Scoping Review, Journal Of Medical Internet Research, vol.24, 2022.
- [19] Faezeh Marzbanrad, Ahsan H. Khandoker, Brett D. Hambly, "Methodological Comparisons of Heart Rate Variability Analysis in Patients With Type 2 Diabetes and Angiotensin Converting Enzyme Polymorphism", IEEE Journal Of Biomedical And Health Informatics, Vol. 20, No. 1, January 2016.
- [20] Konstantia Zarkogianni, Maria Athanasiou, Anastasia C. Thanopoulou, and Konstantina S. Nikita, "Comparison of Machine Learning Approaches Toward Assessing the Risk of Developing Cardiovascular Disease as a Long-Term Diabetes Complication", IEEE Journal Of Biomedical And Health Informatics, VOL. 22, NO. 5, SEPTEMBER 2018.

- [21] Michele Bernardini , Luca Romeo , Paolo Misericordia, and Emanuele Frontoni ,” Discovering the Type 2 Diabetes in Electronic Health Records Using the Sparse Balanced Support Vector Machine”, IEEE Journal Of Biomedical And Health Informatics, Vol. 24, No. 1, January 2020.
- [22] Lambert JR, Perumal E (2022) Oppositional frefy optimization based optimal feature selection in chronic kidney disease classification using deep neural network. J Ambient Intell Humaniz Comput 13(4):1799–1810
- [23] Hajitha shemi V, Hassani Z, Dehmajnoonie I, Borna K (2019), Hybrid algorithms of whale optimization algorithm and k-nearest neighbor to predict the liver disease. EAI Endorsed Trans Con text-Aware Syst Appl 6(16):156838. <https://doi.org/10.4108/eai.13-7-2018.156838>
- [24] Gautam R, Kaur P, Sharma M (2019) A comprehensive review on nature inspired computing algorithms for the diagnosis of chronic disorders in human beings. Prog Artif Intell 8(4):401–424. <https://doi.org/10.1007/s13748-019-00191-1>
- [25] Kaur P, Sharma M (2019) Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: a meta-analysis. J Med Syst. <https://doi.org/10.1007/s10916-019-1341-2>.
- [26] Sharma M, Kaur P (2020) A comprehensive analysis of nature inspired meta-heuristic techniques for feature selection problem. Arch Computat Methods Eng. <https://doi.org/10.1007/s11831-020-09412-6>
- [27] Chander S, Padmanabha V, Mani J (2021) Jaya spider monkey optimization-driven deep convolutional LSTM for the prediction of COVID’19. Bio-Algorithms Med-Syst. <https://doi.org/10.1515/bams-2020-0030>.