https://doi.org/10.33472/AFJBS.6.6.2024.7074-7086

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A Novel Learning on Gestation Diabetic by Hot Deck Imputation and Without Imputation Methods

T.Sujatha[1], Dr.K.R.Ananthapadmanaban[2]

[1]Research Scholar *Department of Computer Science, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India*

[2]Professor & Dean, *Department of Computer Science, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India*

Email: [1]tsujathasrm@gmail.com, [2]kranantp@srmist.edu.in

Article Info

Volume 6, Issue 6, July 2024

Received: 14 May 2024

Accepted: 22 June 2024

Published: 11 July 2024

doi: 10.33472/AFJBS.6.6.2024.7074-7086

ABSTRACT:

Pregnant women all over the globe face the serious health risk of gestational diabetes mellitus (GDM). A type of glucose intolerance, it is characterized by the development or discovery of high blood sugar levels during pregnancy. Not only does gestational diabetes mellitus (GDM) affect the mother's health, but it also puts the unborn child at risk of complications like macrosomia, birth defects, and the need for a cesarean section. The aforementioned issues can be addressed with the aid of machine learning. Hot Deck Imputation consistently outperforms No Imputation across all models tested (Bayes Net, Decision Table, IBK, Multi-Layer Perceptron, and Random Forest). The accuracy for Hot Deck Imputation ranges from 95.97% to 96.97%, while No Imputation accuracy ranges from 78.33% to 83.27%. This substantial difference is also reflected in other metrics such as Precision, Recall, ROC, and PRC, where Hot Deck Imputation shows higher values. The MLP model with Hot Deck Imputation achieves the highest accuracy at 96.43%, though it also has the longest processing time at 7.75 seconds. In contrast, the IBK model with Hot Deck Imputation offers a good balance of high accuracy (96.97%) and the fastest processing time (0.01 seconds). Overall, these results strongly suggest that Hot Deck Imputation significantly improves model performance across various algorithms compared to using datasets with missing values (No Imputation).

Keywords: IBK, Hot Deck Imputation, Bayes Net, Decision Table, Multi-Layer Perceptron

1. Introduction

Variations in maternal age, ethnicity, and body mass index (BMI) contribute to the worldwide variation in the prevalence of gestational diabetes mellitus (GDM). Gestational diabetes mellitus prevalence varies by demographic and geographical area, impacting approximately 7% of pregnancies globally. Along with the alarming rise in obesity and sedentary lifestyles, the incidence of GDM has been steadily climbing over the past few years.

In order to reduce threats to the health of both the mother and the unborn child, it is essential to recognize gestational diabetes at an early stage and to manage the condition effectively. Glucose tolerance tests, usually administered in the second trimester of pregnancy, are the gold standard for diagnosing gestational diabetes mellitus type 1. On the other hand, new studies are looking at how machine learning models can predict GDM, which could lead to better and faster ways to identify people at risk.

Objective:

The following four areas could potentially guide this research into GDM prediction with ML models.

Evaluate the efficacy of different machine learning models in predicting gestational diabetes mellitus (GDM) during pregnancy. These models include K-nearest neighbors (KNN), decision trees, random forests, and multi-layer perceptrons. With this goal in mind, we will evaluate each model by comparing their respective accuracy, recall, ROC, precision-recall, kappa, F, Matthews Correlation Coefficient, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Relative Root Square Error (RRSE).

Optimal Model Identification: Find the machine learning model that reliably and accurately predicts GDM. In order to improve early detection and intervention for gestational diabetes mellitus (GDM) during pregnancy, this objective seeks to determine the model that produces the most accurate and consistent predictions.

Discover what factors influence GDM prediction. Determining how well machine learning models predict GDM depends on a number of parameters, including maternal demographics (age, ethnicity), maternal health indicators (BMI), and lifestyle factors. In order to develop more tailored strategies for evaluating and managing the risks associated with GDM, it is important to understand how these factors affect the accuracy of predictions.

Recommend the best machine learning model for GDM prediction in order to influence clinical practice by turning research findings into practical insights. To achieve this goal, we must enhance maternal and fetal health outcomes by assisting healthcare practitioners in making evidence-based decisions regarding the use of predictive models to identify and treat gestational diabetes mellitus (GDM) during pregnancy. In the end, this study's results could change clinical practice by revealing the best GDM predictive model, which would allow for early diagnosis and treatment to boost fetal and maternal outcomes.

The following is how the rest of the article is structured: Segment 2 depicts related works, whereas Segment 3 depicts materials and procedures. The discussions and results are accessible in Segment 4; lastly, in section 5, the conclusion is presented.

2. Literature Survey

Several studies have shown that ML-based models can be useful for predicting the risk of GDM and for clinical decision-making. These models use a variety of ML techniques, including as logistic regression, decision trees, and support vector machines, all of which have their advantages in terms of interpretability, scalability, and prediction performance. Notably, combining ML with other computational methods, such as genetic algorithms or deep learning, can make GDM prediction models even more accurate and widely used.[1-5] Despite the benefits, ML has not yet been applied to the prediction and management of GDM due to concerns over data quality, the interpretability of models, and clinical implementation. Thoroughly investigating ethical considerations, privacy concerns, and legal constraints is also essential before implementing ML-based healthcare solutions. However, there is significant hope that ML in conjunction with clinical expertise can revolutionize the treatment of GDM, paving the way for more targeted and personalized therapy that meets the specific requirements of each patient.

By detecting subtle patterns in maternal health data, ML approaches can aid in the early detection of gestational diabetes mellitus (GDM), which may exist before the condition manifests clinically. the years 2011–2014 Nutritional counseling and glucose monitoring are two early treatments that can help limit the dangers of uncontrolled hyperglycemia during pregnancy, which can be made possible by early prediction of GDM. In addition, ML algorithms can enhance personalization of GDM management methods, optimization of treatment regimens, and maternal and fetal outcomes by analyzing patient-specific data.[5-10]. From 2019 to 2021, 489 individuals took part in this research. Prior to participating, patients provided their informed permission. The GD clinical decision support system made use of Bayesian optimization and deep learning. A decision-support model was developed by Bayesian optimization using RNN-LSTM. The model's specificity was 99% and sensitivity was 95% when it came to identifying patients at risk of GD. The area under the curve (AUC) for the model was 0.95 with a 95% confidence interval of 1.00. 0.001 is the p-value. In order to save time, cut expenses, and minimize side effects, the clinical diagnosis procedure avoids doing OGTT that is not necessary for patients who do not have a risk of GD [9]. A birth cohort with several features was used to evaluate AI models for GDM prediction. A possible etiology for gestational diabetes mellitus was explored in the research [10]. Diabetic machine learning approaches were deemed inadequate in this investigation. Predictive models for diabetes using machine learning were covered in this article. Clinical decision support for diabetes using machine learning was one of the many medical and technological concerns tackled in this study [11]. By utilizing robust machine learning methods, the GDM Predictor endeavors to foretell the occurrence of GDM in pregnant women. Biochemical markers including A1MG, TBA, BMG, CysC FPG, CO2, and CREA are also considered while designing a patient's treatment plan. [12]. Through the utilization of early trimester patient data and machine learning, Wu et al. developed an algorithm for the prediction of gestational diabetes. Within that range, the model's AUC was between 0.70 and 0.77. A smartphone app was able to detect gestational diabetes by utilizing conventional machine learning techniques. The program made use of information from 12,304 pregnant women [13]. A New-Stacking model was created in this study, and it has the potential to beat current models in terms of ROC, specificity, and accuracy. On the other hand, the SVM model was suggested for clinical prediction because to its high sensitivity [14]. The study discovered that placental dysfunction may impact fetal growth and development in gestational diabetes mellitus. Since there are no effective treatments for GDM and LGA at the moment, it may be worthwhile to investigate pathway inhibitors associated with factor alterations [15]. Algorithms developed for predicting infant development accurately identify gestational diabetes mellitus in pregnant mothers during the first trimester. Body composition and risk variables are part of the model. In order to identify high-risk pregnancies and refer moms for specialized care, hospitals may employ this technique [16]. Of the 735 women surveyed, 190 were expecting a child. Pregnant women were included in both the test and training groups. An area under the curve (AUC) of 0.946 and a prediction accuracy of 0.875 were achieved by the 20-predictor XG Boost ML model. The standard logistic regression (LR) model, on the other hand, relied on just four variables. Prediction accuracy was 0.786 and area under the curve was 0.752. In order to calibrate both variants, the HL test and calibration tables are utilized. According to DCA, the XG Boost ML model prioritizes treating high-risk

GDM women over all other women, or none at all. When it came to discriminatory prediction, the XG Boost machine learning model was ahead of the logistic regression model [17]. Accurately adjusted the settings on both instruments. Random forest regression was used to forecast GDM. Laboratory indicators, personal and family medical history, and the results of a physical examination were all part of the model. Early gestational diabetes mellitus (GDM) is accurately predicted in pregnant women by the trained prediction model [18]. In order to use the random forest approach to forecast early-stage GDM, you need to provide the positive and negative case distribution criteria in your sample datasets. In comparison to Xinhua Hospital Chongming branch (XHCM), SPNPH had 27,95 pregnancies.

3. Materials and Methods

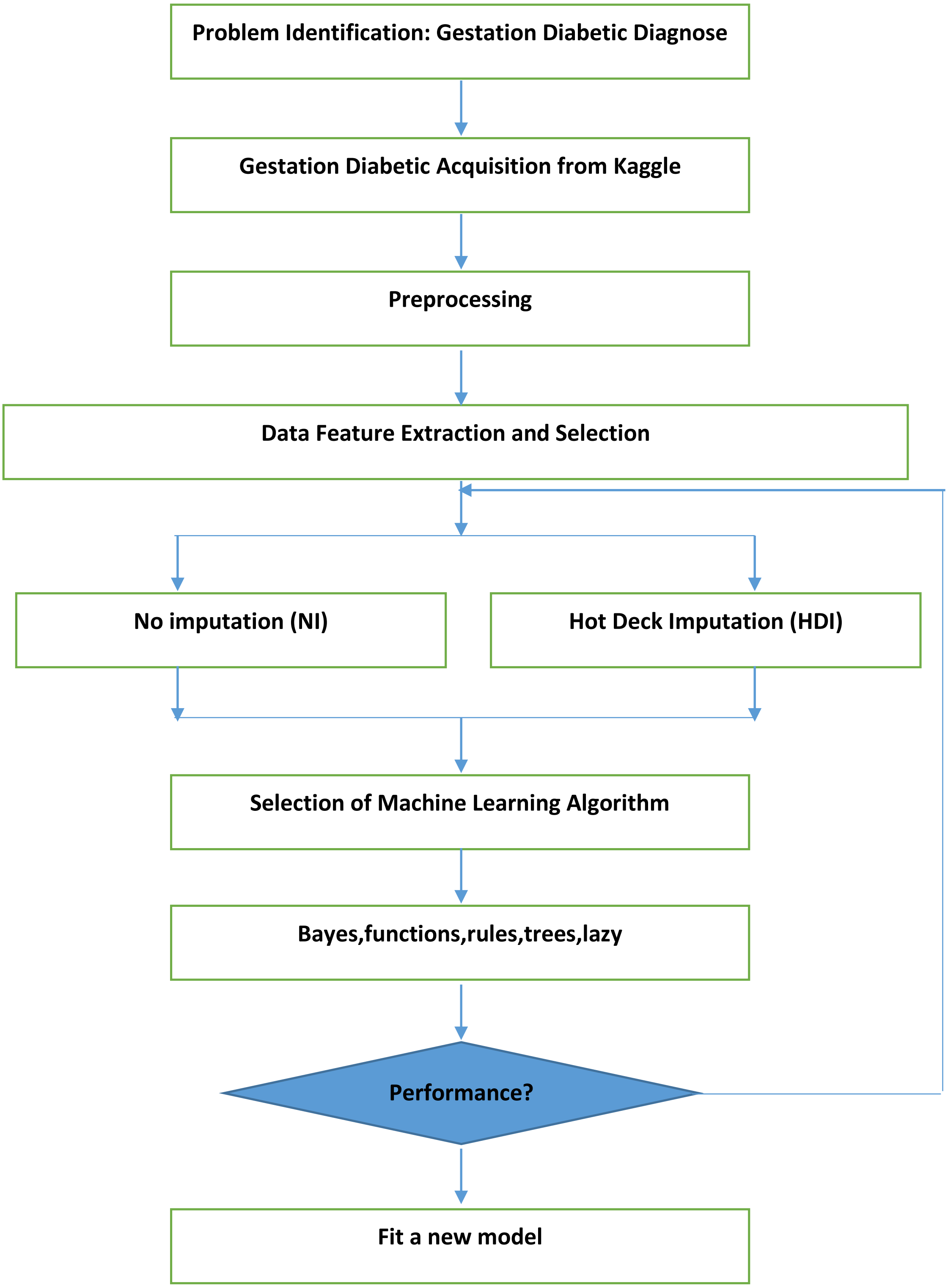
Dataset Collection:
This study analyzes a Kaggle dataset on Gestational Diabetes Mellitus (GDM). The collection contains 3525 cases with 17 variables that provide relevant GDM risk factor information. The GDM dataset meta data follows:

- Case Number: Patient ID, a numeric value.
- Patient Age: Ranging from 20 to 45 years, displayed as a number.
- The number of pregnancies is specified as {0, 1, 2, 3, 4} as a numeric value.
- Previous pregnancy gestation: categorized as {0, 1, 2}, represented as a numeric value.
- The Body Mass Index (BMI) ranges from 13.3 to 45 as a quantitative value.
- High-Density Lipoprotein levels (HDL) range from 15 to 70, expressed numerically.
- Family History: Diabetes history (0=No, 1=Yes), represented as a numeric value.
- The presence of unexplained prenatal loss is indicated by a numeric value (0=No, 1=Yes).
- The history of huge child or birth default is classed as {0=No, 1=Yes} and expressed as a number value.
- Polycystic Ovary Syndrome (PCOS) is categorized as {0=No, 1=Yes} as a numeric number.
- The Systolic Blood Pressure (Sys BP) ranges from 90 to 185 and is represented numerically.
- Dia BP: Numeric value representing Diastolic Blood Pressure (60-124).
- Oral Glucose Tolerance Test (OGTT) results (80-403) are reported as numeric values.
- The hemoglobin levels range from 8.8 to 18 and are expressed as a numeric value.
- The presence of a sedentary lifestyle is indicated by a numeric number (0=No, 1=Yes).
- Presence of prediabetes is categorized as {0=No, 1=Yes} and given as a numeric value.
- Result (GDM/Non GDM): A string representing the outcome (0=Non GDM, 1=GDM).

Despite its initial classification as a numeric data type, the outcome variable in this study—which determines whether the individual has Gestational Diabetes Mellitus (GDM) or not—is treated as a string variable. One possible explanation is that the outcome variable is categorical, with 0 indicating no GDM and 1 indicating the existence of GDM.

4. Methods

Efficient strategies were implemented in Weka 3.9.5, one of the top open-source software, to achieve optimal results. [19-28] This study utilizes a mere 10% of the complete dataset and employs 10-fold cross validation for all categories.



No Yes

Figure 1: Proposed System

5. Results and Discussions

This section will now delve into the results and discussion of this work. Utilizing ML algorithms are employed to achieve optimal results. The abbreviations represent various models: BN (Bayesian Network), DT (Decision Tree), IBK (Instance-Based Learning), MLP (Multi-Layer Perceptron), and RF (Random Forest).

Table 1: Performance of selected Machine Learning Algorithms by HDI Vs NI

S.No	Category	Model	Accuracy	Precision	Recall	ROC	PRC	Time
1	Hot Deck Imputation	BN	96.91%	0.97	0.97	0.99	0.99	0.09
2	No imputation	BN	82.13%	0.89	0.82	0.88	0.88	0.13
3	Hot Deck Imputation	DT	95.97%	0.96	0.96	0.99	0.99	0.5
4	No imputation	DT	82.70%	0.83	0.83	0.89	0.88	0.41
5	Hot Deck Imputation	IBK	96.97%	0.97	0.97	0.97	0.96	0.01
6	No imputation	IBK	83.27%	0.83	0.83	0.75	0.8	0
7	Hot Deck Imputation	MLP	96.43%	0.96	0.96	0.98	0.99	7.75
8	No imputation	MLP	81.18%	0.81	0.81	0.85	0.87	1.34
9	Hot Deck Imputation	RF	96.82%	0.97	0.97	0.97	0.99	1.16
10	No imputation	RF	78.33%	0.79	0.79	0.87	0.87	0.58

The above Table 1 shows that the performance of selected machine learning algorithm by using HDI and NI methods.

HDI with BN model has accuracy (96.91%), precision (0.97), recall (0.99), ROC (0.99), PRC (0.99), time complexity (0.09 seconds); NI with DT model has accuracy (82.13%), precision (0.89), recall (0.82), ROC (0.88), PRC (0.88), time complexity (0.13 seconds);HDI with DT model has accuracy (95.97%), precision (0.96), recall (0.96), ROC (0.99), PRC (0.99), time complexity (0.5 seconds); NI with DT model has accuracy (82.70%), precision (0.83), recall (0.83), ROC (0.89), PRC (0.88), time complexity (0.41 seconds);HDI with IBK model has accuracy (96.97%), precision (0.97), recall (0.97), ROC (0.97), PRC (0.96), time complexity (0.01 seconds); NI with IBK model has accuracy (83.27%), precision (0.83), recall (0.83), ROC (0.75), PRC (0.80), time complexity (0 seconds);HDI with MLP model has accuracy (96.43%), precision (0.96), recall (0.96), ROC (0.98), PRC (0.99), time complexity (7.75 seconds); NI with MLP model has accuracy (81.18%), precision (0.81), recall (0.81), ROC (0.85), PRC (0.87), time complexity (1.34 seconds);HDI with MLP model has accuracy (96.82%), precision (0.97), recall (0.97), ROC (0.97), PRC (0.99), time complexity (1.16 seconds); NI with MLP model has accuracy (78.33%), precision (0.79), recall (0.79), ROC (0.87), PRC (0.87), time complexity (0.58 seconds).

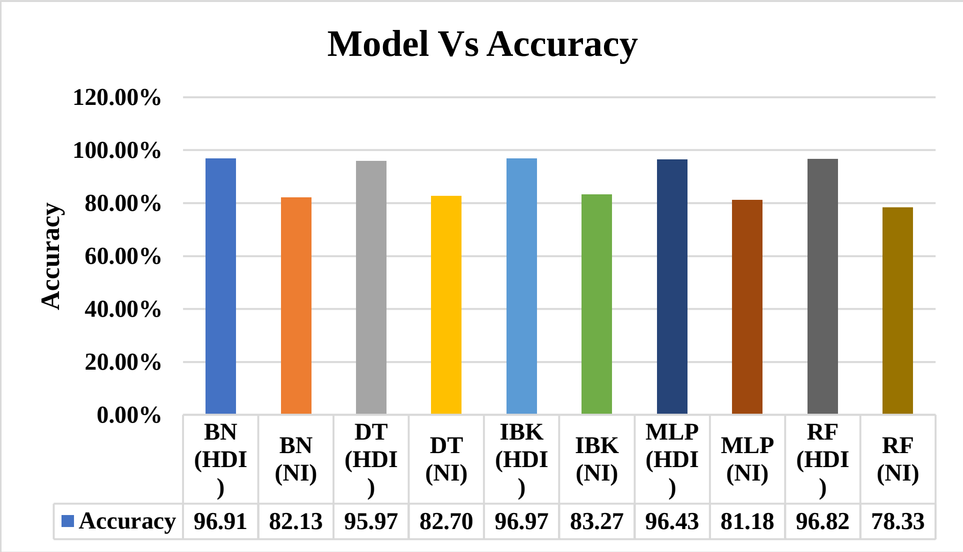


Figure 2: Accuracy performance of selected learning by by HDI Vs NI methods

The above diagram 2 shows that the performance of accuracy on selected ML models by HDI and NI methods.

This bar graph compares the accuracy of five different machine learning models (BN, DT, IBK, MLP, and RF) under two conditions: Hot Deck Imputation (HDI) and No Imputation (NI). For each model, the accuracy with Hot Deck Imputation (HDI) is noticeably higher than with No Imputation (NI).All models using HDI achieve accuracies above 95%, with IBK showing the highest at 96.97%.Without imputation, accuracies range from about 78% to 83%, significantly lower than their HDI counterparts.The Random Forest (RF) model shows the most dramatic difference between HDI (96.82%) and NI (78.33%), an improvement of over 18 percentage points.The Decision Tree (DT) model has the smallest difference between HDI (95.97%) and NI (82.70%), though it's still a substantial improvement of over 13 percentage points.The pattern of improvement with HDI is consistent across all five models, suggesting that Hot Deck Imputation is effective regardless of the specific machine learning algorithm used. The IBK model with HDI shows the highest accuracy at 96.97%, closely followed by BN and RF with HDI. This visualization clearly demonstrates the significant positive impact of Hot Deck Imputation on model accuracy across various machine learning algorithms, reinforcing the importance of addressing missing data in improving model performance.

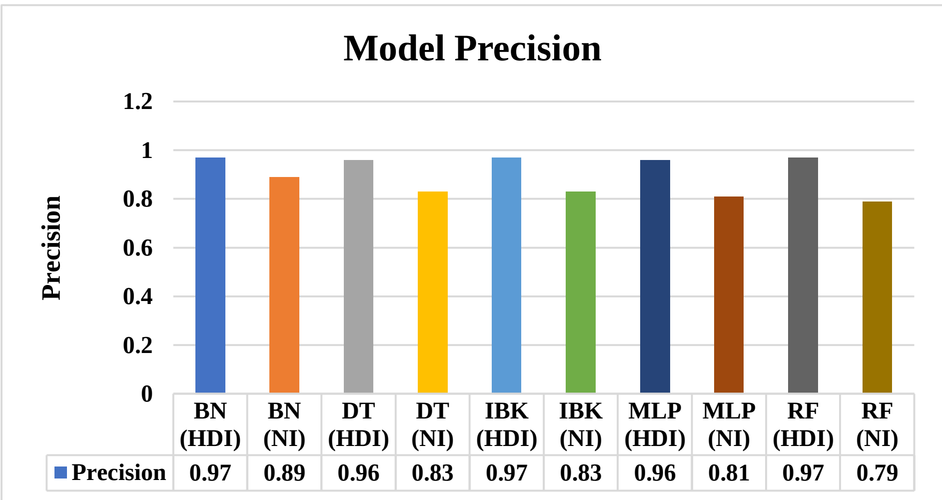


Figure 3: Precision performance of selected learning by by HDI Vs NI methods

The above diagram 4 shows that the performance of precision on selected ML models by HDI and NI methods. This bar graph compares the precision of five different machine learning models (BN, DT, IBK, MLP, and RF) under two conditions: Hot Deck Imputation (HDI) and No Imputation (NI). For each model, the precision with Hot Deck Imputation (HDI) is higher than with No Imputation (NI). Models using HDI achieve precision scores of 0.96 or 0.97, indicating very high precision.Without imputation, precision scores range from 0.79 to 0.89, which are noticeably lower than their HDI counterparts. The Random Forest (RF) model shows the most significant difference between HDI (0.97) and NI (0.79), an improvement of 0.18.The Bayesian Network (BN) model has the smallest difference between HDI (0.97) and NI (0.89), though it's still a notable improvement of 0.08.The pattern of improvement with HDI is consistent across all five models, suggesting that Hot Deck Imputation effectively enhances precision regardless of the specific algorithm used.BN, IBK, and RF models with HDI all achieve the highest precision score of 0.97. This visualization clearly demonstrates that Hot Deck Imputation significantly improves model precision across various machine learning algorithms. Higher precision indicates that when these models predict a positive class, they are more likely to be correct, which is crucial in many real-world applications where false positives can be costly or problematic.

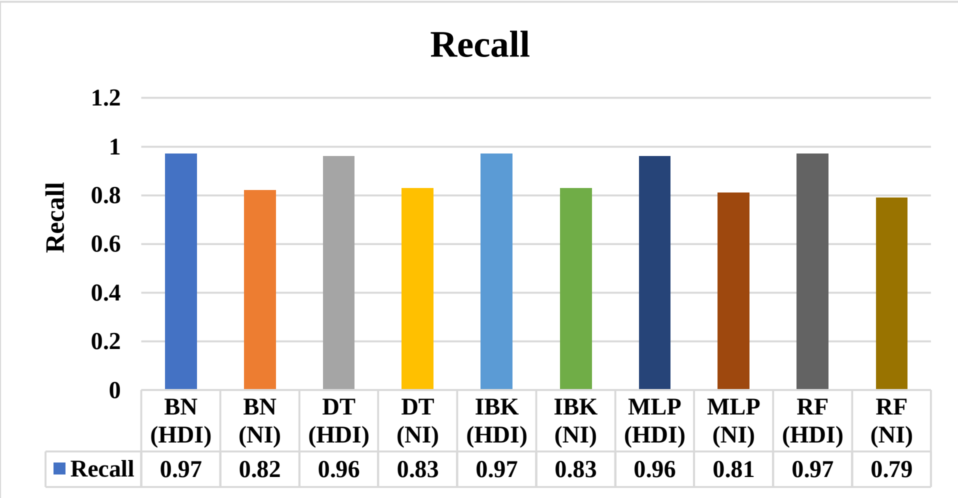


Figure 4: Recall performance of selected learning by by HDI Vs NI methods

The above diagram 4 shows that the performance of recall on selected ML models by HDI and NI methods. This bar graph compares the recall of five different machine learning models (BN, DT, IBK, MLP, and RF) under two conditions: Hot Deck Imputation (HDI) and No Imputation (NI). For each model, the recall with Hot Deck Imputation (HDI) is higher than with No Imputation (NI). Models using HDI achieve recall scores of 0.96 or 0.97, indicating very high precision. Without imputation, recall scores range from 0.79 to 0.83, which are noticeably lower than their HDI counterparts. The Random Forest (RF) model shows the most significant difference between HDI (0.97) and NI (0.79), an improvement of 0.18. The IBK (BN) model has the smallest difference between HDI (0.97) and NI (0.83). The pattern of improvement with HDI is consistent across all five models, suggesting that Hot Deck Imputation effectively enhances precision regardless of the specific algorithm used.BN, IBK, and RF models with HDI all achieve the highest precision score of 0.97. This visualization clearly demonstrates that Hot Deck Imputation significantly improves model precision across various machine learning algorithms. Higher precision indicates that when these models predict a positive class, they are more likely to be correct, which is crucial in many real-world applications where false positives can be costly or problematic.

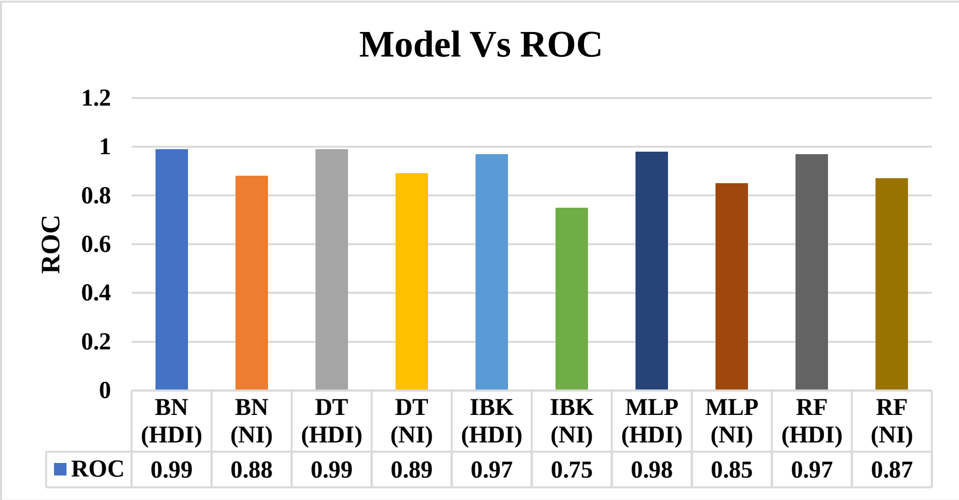


Figure 5: ROC performance of selected learning by by HDI Vs NI methods

The above diagram 5 shows that the Performance of ROC levels on selected ML models by HDI and NI methods. This bar graph compares the ROC (Receiver Operating Characteristic) scores of five different machine learning models (BN, DT, IBK, MLP, and RF) under two conditions: Hot Deck Imputation (HDI) and No Imputation (NI). For each model, the ROC score with Hot Deck Imputation (HDI) is higher than with No Imputation (NI). Models using HDI achieve very high ROC scores, ranging from 0.97 to 0.99. This indicates excellent discriminative ability. Without imputation, ROC scores range from 0.75 to 0.89, which are noticeably lower than their HDI counterparts. The IBK model shows the most significant difference between HDI (0.97) and NI (0.75), an improvement of 0.22. The BN model has the smallest difference between HDI (0.99) and NI (0.88), though it's still a notable improvement of 0.11.The pattern of improvement with HDI is consistent across all five models, suggesting that Hot Deck Imputation effectively enhances the ROC score regardless of the specific algorithm used.BN and DT models with HDI achieve the highest ROC score of 0.99, closely followed by MLP at 0.98.This visualization clearly demonstrates that Hot Deck Imputation significantly improves the ROC scores across various machine learning algorithms. Higher ROC scores indicate better model performance in distinguishing between classes, regardless of the chosen classification threshold. This improvement in discriminative ability is crucial for creating more reliable and effective predictive models.

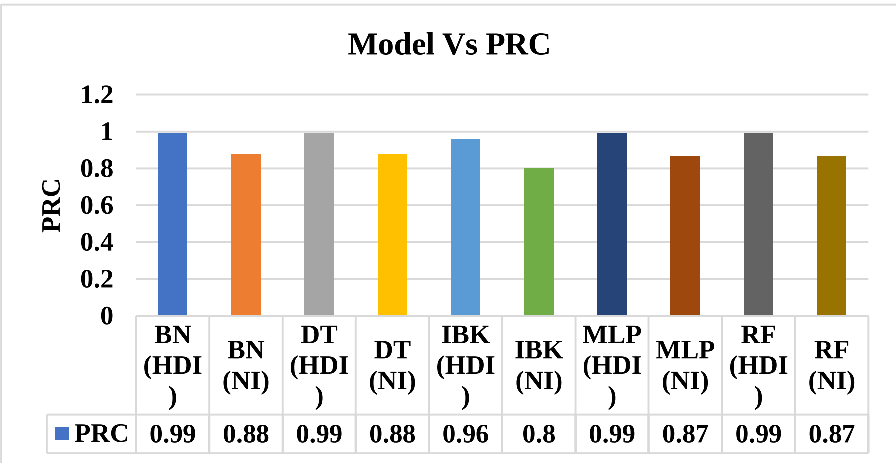


Figure 6: PRC performance of selected learning by by HDI Vs NI methods

The above diagram 6 shows that the Performance of PRC levels on selected ML models by HDI and NI methods. This bar graph compares the PRC (Precision-Recall Curve) scores of five different machine learning models (BN, DT, IBK, MLP, and RF) under two conditions: Hot Deck Imputation (HDI) and No Imputation (NI). For each model, the PRC score with Hot Deck Imputation (HDI) is higher than with No Imputation (NI). The Models using HDI achieve very high PRC scores, ranging from 0.96 to 0.99. This indicates excellent performance in balancing precision and recall. Without imputation, PRC scores range from 0.80 to 0.88, which are noticeably lower than their HDI counterparts. The IBK model shows the most significant difference between HDI (0.96) and NI (0.80), an improvement of 0.16.The DT model has the smallest difference between HDI (0.99) and NI (0.88), though it's still a notable improvement of 0.11.The pattern of improvement with HDI is consistent across all five models, suggesting that Hot Deck Imputation effectively enhances the PRC score regardless of the specific algorithm used.BN, DT, MLP, and RF models with HDI all achieve the highest PRC score of 0.99, with IBK following closely at 0.96. This visualization clearly demonstrates that Hot Deck Imputation significantly improves the PRC scores across various machine learning algorithms. Higher PRC scores indicate better model performance in maintaining both high precision and high recall, which is particularly important in imbalanced classification problems or when the cost of false positives and false negatives differs. The consistent improvement across all models underscores the effectiveness of Hot Deck Imputation in enhancing overall model

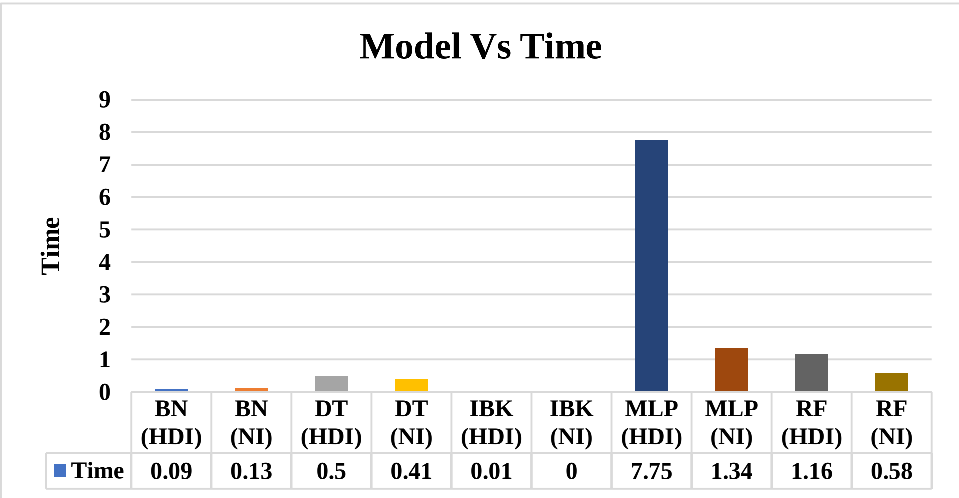


Figure 7: Time Consumption performance of selected learning by HDI Vs NI methods

The above diagram 7 shows that the Performance of Time Consumptions to construct models on selected ML models by HDI and NI methods. The bar chart compares different machine learning models based on their processing time. The tallest bar corresponds to MLP (NI), indicating that this model takes significantly more time than others. The BN (HDI) model and BN (NI) model have similar processing times. The DT (HDI) model and DT (NI) model also show similar performance. The IBK (HDI) model and IBK (NI) model have comparable processing times. The RF (HDI) model and RF (NI) model exhibit similar behavior.

6. Conclusions

The data compares the performance of five different machine learning models (BN, DT, IBK, MLP, and RF) under two conditions: using Hot Deck Imputation to handle missing data, and using No Imputation (i.e., leaving missing values as is). Hot Deck Imputation shows a clear and consistent advantage across all models and metrics. The accuracy improvements are substantial, with increases ranging from 13.28 percentage points (for DT) to 18.49 percentage

points (for RF). This indicates that Hot Deck Imputation is highly effective in improving model performance, likely by preserving important data relationships and distributions.
Research work finds :

- BN (Bayesian Network): Shows a 14.78% improvement in accuracy with imputation.
- DT (Decision Tree): Accuracy increases by 13.27% with imputation.
- IBK (Instance-Based K-nearest neighbor): Demonstrates a 13.70% accuracy boost.
- MLP (Multi-Layer Perceptron): Exhibits a 15.25% increase in accuracy.
- RF (Random Forest): Shows the most dramatic improvement, with an 18.49% increase in accuracy.

Beyond accuracy, other metrics (Precision, Recall, ROC, PRC) consistently show better performance with Hot Deck Imputation. This suggests that the imputation method not only improves overall accuracy but also enhances the models' ability to correctly identify positive cases (precision) and capture all positive cases (recall).
The time taken for model processing varies, with MLP taking significantly longer than other models, especially with imputation. The IBK model with Hot Deck Imputation stands out for its combination of high accuracy (96.97%) and extremely fast processing time (0.01 seconds). These results strongly support the use of Hot Deck Imputation as a preprocessing step in machine learning pipelines, especially when dealing with datasets that have missing values. It appears to be a robust method that can significantly enhance model performance across various algorithms and evaluation metrics.

7. References

1. Sumathi A, S. Meganathan and B. Vijila Ravisankar, "An intelligent gestational diabetes diagnosis model using deep stacked autoencoder," Computers, Materials & Continua, vol. 69, no.3, pp. 3109–3126, 2021.https://www.techscience.com/cmc/v69n3/44131.
2. Yipeng Wang, Pei Sun, Zicheng Zhao, Yousheng Yan, Wentao Yue, Kai Yang, Ruixia Liu, Hui Huang, Yinan Wang, Yin Chen, Nan Li, Hailong Feng, Jing Li, Yifan Liu, Yujiao Chen, Bairong Shen, Lijian Zhao, Chenghong Yin, Identify gestational diabetes mellitus by deep learning model from cell-free DNA at the early gestation stage, Briefings in Bioinformatics, Volume 25, Issue 1, January 2024, bbad492, https://doi.org/10.1093/bib/bbad492
3. Varada Vivek Khanna, Krishnaraj Chadaga, Niranjana Sampathila, Srikanth Prabhu, Rajagopala Chadaga P., Devadas Bhat & Swathi K. S. (2024) Explainable artificial intelligence-driven gestational diabetes mellitus prediction using clinical and laboratory markers, Cogent Engineering, 11:1, 2330266, DOI: 10.1080/23311916.2024.2330266
4. Palanivinayagam, A.; Damaševičius, R. Effective Handling of Missing Values in Datasets for Classification Using Machine Learning Methods. Information 2023, 14, 92. https://doi.org/10.3390/info14020092
5. Li, Yx., Liu, Yc., Wang, M. et al. Prediction of gestational diabetes mellitus at the first trimester: machine-learning algorithms. Arch Gynecol Obstet (2023). https://doi.org/10.1007/s00404-023-07131-4
6. Kurt, B., Gürlek, B., Keskin, S. et al. Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques. Med Biol Eng Comput 61, 1649–1660 (2023). https://doi.org/10.1007/s11517-023-02800-7
7. Yipeng Wang, Pei Sun, Zicheng Zhao, Yousheng Yan, Wentao Yue, Kai Yang, Ruixia Liu, Hui Huang, Yinan Wang, Yin Chen, Nan Li, Hailong Feng, Jing Li, Yifan Liu, Yujiao Chen, Bairong Shen, Lijian Zhao, Chenghong Yin, Identify gestational diabetes mellitus by deep learning model from cell-free DNA at the early gestation stage, Briefings

in Bioinformatics, Volume 25, Issue 1, January 2024, bbad492, https://doi.org/10.1093/bib/bbad492

8. Shahin, M., Chen, F.F., Hosseinzadeh, A. et al. Using machine learning and deep learning algorithms for downtime minimization in manufacturing systems: an early failure detection diagnostic service. Int J Adv Manuf Technol 128, 3857–3883 (2023). https://doi.org/10.1007/s00170-023-12020-w
9. Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, Jiang K, Liu Y, Wu H. Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis. J Med Internet Res. 2022 Mar 16;24(3):e26634. doi: 10.2196/26634. PMID: 35294369; PMCID: PMC8968560.
10. JOUR,Costa, Gianni,Sofany, Hosam,Seoud, Samir A.,Karam, Omar H.Abd El-Latif, Yasser M.,Taj-Eddin, Islam A. T. F. A Proposed Technique Using Machine Learning for the Prediction of Diabetes Disease through a Mobile App, International Journal of Intelligent Systems, 2024,, 0884-8173,Hindawi, https://doi.org/10.1155/2024/6688934.
11. Tuppad, A., Patil, S.D. Machine learning for diabetes clinical decision support: a review. Adv. in Comp. Int. 2, 22 (2022). https://doi.org/10.1007/s43674-022-00034-y
12. Xing, J., Dong, K., Liu, X. et al. Enhancing gestational diabetes mellitus risk assessment and treatment through GDMPredictor: a machine learning approach. J Endocrinol Invest (2024). https://doi.org/10.1007/s40618-024-02328-z
13. Wu YT, Zhang CJ, Mol BW, et al. Early Prediction of Gestational Diabetes Mellitus in the Chinese Population via Advanced Machine Learning. The Journal of Clinical Endocrinology and Metabolism. 2021 Mar;106(3):e1191-e1205. DOI: 10.1210/clinem/dgaa899. PMID: 33351102; PMCID: PMC7947802.
14. Wang, J., Chen, X., Pan, Y., Chen, K., Zhang, Y., Li, Q., … & Wei, L. (2021). Machine learning approaches for early prediction of gestational diabetes mellitus based on prospective cohort study.. https://doi.org/10.21203/rs.3.rs-508626/v1
15. Assi E, D'Addio F, Mandò C, Maestroni A, Loretelli C, Ben Nasr M, Usuelli V, Abdelsalam A, Seelam AJ, Pastore I, Magagnotti C, Abdi R, El Essawy B, Folli F, Corradi D, Zuccotti G, Cetin I, Fiorina P. Placental proteome abnormalities in women with gestational diabetes and large-for-gestational-age newborns. BMJ Open Diabetes Res Care. 2020 Nov;8(2):e001586. doi: 10.1136/bmjdrc-2020-001586. PMID: 33188009; PMCID: PMC7668299.
16. Cremona A, O'Gorman CS, Ismail KI, Hayes K, Donnelly AE, Hamilton J, Cotter A. A risk-prediction model using parameters of maternal body composition to identify gestational diabetes mellitus in early pregnancy. Clin Nutr ESPEN. 2021 Oct;45:312-321. doi: 10.1016/j.clnesp.2021.08.002. Epub 2021 Aug 18. PMID: 34620334.
17. Hu X, Hu X, Yu Y and Wang J (2023) Prediction model for gestational diabetes mellitus using the XG Boost machine learning algorithm. Front. Endocrinol. 14:1105062. doi: 10.3389/fendo.2023.1105062.
18. Wei LL, Pan YS, Zhang Y, Chen K, Wang HY, Wang JY. Application of machine learning algorithm for predicting gestational diabetes mellitus in early pregnancy. Front Nurs. 2021;3:209–222.
19. Ayyappan, G., Venkatesh Babu, R., Alice, K., Kumaravel, A., Knowledge on Visual representation of binary classification: Apple Vs Avocado, Indian Journal of Computer Science and Engineering, 2022, 13(6), pp. 2035–2046
20. Venkatesh Babu, R., Ayyappan, G., Kumaravel, A., Novel ensemble approaches to model macroscopic material behavior using micromechanical simulations,International Journal of Scientific and Technology Research, 2019, 8(11), pp. 2560–2564
21. Begum, A., David, S.A., Hemalatha, D. and Kollipara, L.S.S., 2023, May. Deep Learning-Based Lung Cancer Classification: Recent Developments and Future Prospects.

In 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-8). IEEE.

22. Begum, A., Gattamaneni, N., Mekala, S.K. and Junaidul, M.D., 2023, May. Medical Data Governance System using Blockchain. In 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-6). IEEE.
23. Senthil, K., Begum, A., Ayyappan, G. and Sundaravadivazhagan, B., 2024. Artificial Intelligence and Automation in the Transport Industry: Envisioning the Future of Mobility. In Intelligent Systems and Industrial Internet of Things for Sustainable Development (pp. 235-265). Chapman and Hall/CRC.
24. Begum, A., Dhilip Kumar, V., Asghar, J., Hemalatha, D. and Arulkumaran, G., 2022. A combined deep CNN: LSTM with a random forest approach for breast cancer diagnosis. Complexity, 2022(1), p.9299621.
25. Kavitha, P., Ayyappan, G., Jayagopal, P., ...Alqahtani, M.S., Soufiene, B.O. Detection for melanoma skin cancer through ACCF, BPPF, and CLF techniques with machine learning approach, BMC Bioinformatics, 2023, 24(1), 458.
26. Venkatesh Babu, R., Ayyappan, G., Kumaravel, A., Exploratory data analysis on macroscopic material behavior using micromechanical simulations by applying the gaussian processes with various kernels, Indian Journal of Computer Science and Engineering, 2021, 12(1), pp. 246–253
27. Ayyappan, G., Alice Suresh, K., Thirunavukkarasu, S., Kumaravel, A.,Construction of fuzzy inference system for public food supply chain in cooperative sector, Indian Journal of Computer Science and Engineering, 2020, 11(6), pp. 948–952.
28. Srinivasan S., Ananthapadmanaban K.R. (2020) Intelligent Agent-Based Organization for Studying the Big Five Personality Traits. In: Peng SL., Son L., Suseendran G., Balaganesh D. (eds) Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, vol 118. Springer, Singapore. https://doi.org/10.1007/978-981-15-3284-9_10