MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 657-663

Early-Stage Diabetes Prediction using Optinet Deep learning techniques considering Comorbid Health Conditions

S. Padmapriya[1], Dr. C. Kavitha[2]
[1]Research Scholar, Department of Computer Science, Thiruvalluvar Govt. Arts College, Rasipuram. 637 401
[2]Associate Professor, Department of Computer Science, Thiruvalluvar Govt. Arts College, Rasipuram. 637 401

Abstract
This research presents an innovative methodology for early diabetes prediction, taking into consideration comorbid health conditions. The approach, named "DiabNetTune," integrates advanced deep learning techniques, including optimized Recurrent Neural Networks (RNNs), alongside Convolutional Neural Networks (CNNs) and traditional Random Forest algorithms. By harnessing RNNs to capture temporal dependencies within clinical data and CNNs to extract spatial features, the model provides a comprehensive analysis of the intricate relationships between clinical variables and diabetes onset. Furthermore, the incorporation of Random Forest enhances predictive accuracy and interpretability, facilitating the identification of key predictive factors. Through rigorous experimentation, the proposed framework demonstrates superior performance compared to Deep learning methods, underscoring the efficacy of optimized RNN hyperparameters, CNNs, and Random Forest in advancing personalized diabetes risk assessment and intervention strategies.

Keywords: Diabetes Prediction, Comorbid Health Conditions, Deep Learning, RNN, CNN, Random Forest

Introduction

Globally, diabetes mellitus—a chronic metabolic disease marked by high blood sugar—poses a serious threat to public health. Effective early identification and treatments are now critical to limiting its related problems and lessening the strain on healthcare systems, as its prevalence continues to rise. However, because diabetes is frequently asymptomatic and is influenced by a variety of genetic, lifestyle, and clinical variables, detecting the disease in its early stages can be extremely difficult.Furthermore, diabetes is seldom isolated; rather, it commonly coexists—a phenomenon known as comorbidity—with other long-term medical disorders. Comorbid conditions including obesity, dyslipidemia, and hypertension not only make managing diabetes more difficult, but they also have a role in the etiology and course of the disease. Thus, understanding the complex interactions between diabetes and co-occurring medical disorders is essential for precise forecasting and individualized treatment plans.To address these issues, this study aims to create a predictive model for the early identification of diabetes that incorporates Optinet deep learning techniques and accounts for the co-occurrence of other medical disorders. The suggested approach seeks to improve diabetes prediction accuracy and reliability by utilizing deep learning techniques and incorporating several clinical data sources, including comorbidity information.

This introduction emphasizes the significance of taking into account comorbid health problems in the creation of predictive models and lays the groundwork for a thorough examination of the complexity involved in diabetes prediction. We hope that our research will contribute to the field of early diabetes identification, opening the door for more focused and efficient therapies that will enhance patient outcomes and lessen the burden of this crippling illness.

Contribution in this research

1. Data integration and preprocessing to handle heterogeneous clinical data sources and comorbidity information.
2. Feature representation and selection techniques that capture the complex interplay between diabetes and comorbid health conditions.
3. Development of robust deep learning architectures, such as Optinet, capable of extracting meaningful patterns and relationships from high-dimensional clinical data.
4. Evaluation of model performance in real-world clinical settings, considering factors such as data variability, patient heterogeneity, and clinical utility.

Literature Review

According to Shaw et al. (2017), diabetes mellitus, which is defined by persistent hyperglycemia, constitutes a huge global health burden with growing prevalence rates all over the world. Diabetes treatment is made more difficult by the presence of comorbid health disorders, such as hypertension, dyslipidemia, and obesity, all of which contribute to the progression and consequences of the illness (Gujral et al., 2017; Zhao et al., 2020). However, the existence of these conditions makes diabetes management even more difficult. It is vital to acknowledge the interconnection of these factors in order to construct efficient prediction models that take into consideration the multidimensional character of the development and course of diabetes.

In light of recent developments in deep learning, there has been a growing interest in the application of complex algorithms for the purpose of improving the accuracy of diabetes prediction. According to Lipton et al. (2015), Recurrent Neural Networks (RNNs) have emerged as a strong tool for processing sequential clinical data. This technology enables the capturing of temporal relationships and longitudinal patterns. In addition, Convolutional Neural Networks (CNNs) have demonstrated their potential in the extraction of spatial information from medical imaging data, which can provide insights into issues connected to diabetes (Esteva et al., 2017). When these deep learning techniques are combined with more conventional machine learning approaches, such as Random Forest, the prediction performance and interpretability of the model are improved, which in turn makes it easier to identify critical risk variables (Breiman, 2001; Christodoulou et al., 2019).

Time-series anomaly detection algorithms have been investigated for the purpose of recognizing early indicators of diabetes development in clinical data (Razavian et al., 2015). These algorithms might be used in conjunction with deep learning processes. Furthermore, gene expression biomarkers have been explored for their potential utility in distinguishing between diabetes subtypes and early-stage diabetes (Natarajan et al., 2017). This is because these biomarkers have the ability to differentiate between normal physiological states and in the early stages of diabetes. In addition, deep learning techniques have been utilized in order to automate the identification of diabetic retinopathy, which is a prevalent condition that has substantial consequences for the outcomes of patients (Oh et al., 2020).

An increase in the amount of research that makes use of machine learning algorithms has been prompted by the need to achieve an accurate and early prediction of diabetes mellitus. A comprehensive assessment of machine learning-based techniques for early diabetes prediction was carried out by Singh and Shukla (2020). The review offered insights into the many methodology and algorithms that are utilized in this field. The findings of their review highlight the necessity of utilizing machine learning techniques in order to construct prediction models that are able to identify individuals who are at risk of developing diabetes over their lifetime.The authors Zhang, Wu, and Zhu (2019) developed an autonomous prediction model for diabetes that makes use of machine learning algorithms. This model highlights the promise of these approaches in helping early diagnosis and intervention. The findings of their research provide a contribution to the expanding body of literature that advocates for the use of machine learning into clinical practice in order to achieve better outcomes associated with healthcare. In a similar vein, Kumar et al. (2019) carried out an exhaustive evaluation of machine learning approaches for the early prediction of diabetes. They placed a strong emphasis on the significance of feature selection and model tuning in terms of improving predictive performance. The results of their investigation offer useful insights into the advantages and disadvantages of various machine learning algorithms when used to the prediction of diabetes.

In 2018, Martínez-Castillo and colleagues conducted an investigation into the categorization of diabetes mellitus by employing machine learning techniques. Their findings demonstrated that these approaches are effective in properly discriminating between persons who have diabetes and those who do not have diabetes. Their research makes a significant contribution to the creation of reliable prediction models for the diagnosis of

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 657-663

diabetes and the classification of risk factors. An improved prediction model for diabetes was suggested by Nagarajan, Suthakar, and Manogaran (2017). This model makes use of useful features, and it emphasizes the significance of feature engineering in terms of enhancing model accuracy. The findings of their study highlight the potential of machine learning to discover meaningful patterns within healthcare data for the purpose of early illness identification and management.The application of deep learning to diabetes prediction has gained significant traction in recent years, with numerous studies exploring various DL architectures and methodologies (Baswaraj 2024). A wide range of DL models have been employed, each with its strengths and limitations. Artificial Neural Networks (ANNs) have been extensively used, demonstrating their capacity to learn complex non-linear relationships between predictor variables and the outcome (diabetes diagnosis) (Baswaraj 2024), (Zhang 2023), Convolutional Neural Networks (CNNs), particularly effective in image processing, have been applied to analyze retinal images for early detection of diabetic retinopathy, a common complication of diabetes (Wang 2025). Recurrent Neural Networks (RNNs), adept at handling sequential data, have been utilized to analyze time-series data such as glucose measurements over time, allowing for the detection of patterns indicative of developing diabetes (Naveed 2023). Long Short-Term Memory (LSTM) networks, a specialized type of RNN, have also been employed to capture long-range dependencies in time-series data. The reported accuracy rates across these studies vary widely, ranging from approximately 68% to an impressive 99.9% (Balaji 2022). This variability reflects differences in datasets, model architectures, preprocessing techniques, and evaluation metrics employed across studies. Critically, many studies focused primarily on individual risk factors and clinical measurements, with limited integration of comorbid health conditions (Zhang 2023). The incorporation of comorbidity data is crucial because the presence of other diseases can significantly influence the development and progression of diabetes, and neglecting this interaction can lead to less accurate predictions. This review highlights the need for a more comprehensive approach that integrates multiple data modalities and explicitly considers the impact of comorbid conditions on diabetes prediction.The literature of this research indicates that there is a growing interest in utilizing machine learning algorithms for the purpose of early diabetes prediction. Studies have highlighted the significance of feature selection, model optimization, and the incorporation of a variety of methodologies in order to improve the accuracy of predictions and the clinical utility of the results.

III. METHODOLOGY

The description of the dataset, the technique that was used (i.e., data preparation, feature ranking, and analysis in terms of the target classes), the risk prediction models, and the evaluation metrics will be the primary areas of emphasis for our research in this section.

Dataset Description

The dataset encompasses a rich repository of clinical data spanning a decade of healthcare interactions across 130 US hospitals and integrated delivery networks, comprising a total of 101,740 records. Encompassing over 50 diverse features, the dataset encapsulates multifaceted insights into patient demographics, hospital outcomes, and medical histories. Notable attributes include patient identifiers, demographic details such as race, gender, and age, admission characteristics such as admission type and duration of hospital stay, as well as clinical specifics such as the medical specialty of the admitting physician, the number of laboratory tests conducted, and HbA1c test results. Additionally, the dataset encompasses a comprehensive array of comorbid conditions and medication-related attributes, with a focus on diabetic medications and outpatient, inpatient, and emergency visit frequencies in the year preceding hospitalization. Of particular significance are the 24 medication-related features, which detail the prescription status and dosage alterations for a range of medications, including metformin, repaglinide, and insulin, among others, providing invaluable insights into medication usage patterns and treatment dynamics.

Data Preprocessing

1. Missing Values:Given the dataset's extensive scope, missing values are likely to be present in demographic attributes, clinical specifics, and medication-related features. Missing BMI values and entries categorized under "Other" gender were removed to maintain data consistency. Imputation techniques were applied to address missing values in comorbid conditions and medication-related attributes, ensuring data completeness for accurate analysis.
2. Encoding: Categorical variables, including patient demographics were transformed using one-hot encoding to enable seamless integration into machine learning models. Medication-related attributes, such as prescription status and dosage changes for drugs like metformin, repaglinide, and insulin, were encoded using one-hot or ordinal encoding, ensuring a structured representation of treatment patterns and medication usage dynamics.
3. Normalization: Continuous variables, such as age, hospital stay duration, number of laboratory tests conducted, and HbA1c test results, were scaled to ensure uniformity across features. This normalization step helps prevent feature dominance, enhances computational efficiency, and improves model performance.
4. Class Imbalance: Given the dataset's focus on patient outcomes and medical histories, class imbalance may be present in factors such as hospitalization frequencies and medication prescription patterns. To address this, the Synthetic Minority Oversampling Technique (SMOTE) was applied to generate synthetic samples for underrepresented classes, ensuring balanced data distribution and improving predictive model reliability.

Problem definition

The challenge addressed in this research work is the early detection of diabetes in persons while taking into consideration the existence of concomitant health problems. Diabetes is a chronic metabolic ailment characterized by high blood sugar levels that, if not treated or identified, can result in major consequences such as cardiovascular disease, renal failure, and blindness. Early identification and action are critical for successful management and avoidance of problems. Comorbid health problems, such as hypertension, obesity, and dyslipidemia, commonly coexist with diabetes, complicating its diagnosis and treatment. These illnesses frequently share risk factors and pathophysiological pathways with diabetes, thus their simultaneous assessment is critical for accurate prognosis and individualized therapeutic options.

Objectives of this research

The purpose of this study is to construct a prediction model that makes use of Optinet deep learning techniques in order to identify individuals who are at risk of acquiring diabetes at an early stage, while also taking into consideration the effect of comorbid health problems. Through the incorporation of a wide range of clinical characteristics and the utilization of the power of deep learning algorithms, the model that has been developed intends to increase the accuracy and reliability of diabetes prediction, which will ultimately lead to the facilitation of prompt intervention and improved patient outcomes.

Diabetes Risk Prediction

Deep learning models are becoming more important in modern healthcare since they allow doctors to automatically evaluate patient risk for disease in different clinical settings. Within this framework, the main objective is to formulate the long-term risk of developing diabetes as a classification job, with two distinct goal classes: 'Diabetes' and 'Non-Diabetes.' After being trained extensively on relevant datasets, these machine-learning models can determine the probability of diabetes incidence based on the values of input characteristics and can predict the class label of unlabeled occurrences. Thorough data preparation, feature extraction using ranking approaches, model training, and performance evaluation are all essential steps in the methodology. Healthcare providers may maximize diabetes risk assessment and intervention techniques by following these methodological principles and utilizing the predictive power of machine learning algorithms.

Information Extraction

The dataset offers a comprehensive reservoir of information encompassing patient demographics, hospital outcomes, medical histories, and medication utilization patterns. Here's an overview of the key data elements that can be extracted and analyzed:

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 657-663

1. Patient Demographics:
Extract details such as race, gender, and age to discern demographic trends and potential variations in healthcare experiences.
2. Admission Characteristics:
Analyze admission types and durations of hospital stays to understand the nature and duration of healthcare interactions.
3. Clinical Details:
Explore the medical specialties of admitting physicians and the frequency of laboratory tests to gain insights into clinical practices and diagnostic procedures.
4. Comorbid Conditions:
Identify comorbid health conditions and their prevalence among patients to assess their impact on healthcare outcomes and treatment approaches.
5. Medication Utilization:
Investigate medication-related features, including prescription statuses and dosage adjustments for diabetic and other medications, to evaluate medication adherence and treatment effectiveness.
6. Healthcare Utilization:
Examine outpatient, inpatient, and emergency visit frequencies in the year before hospitalization to analyze healthcare utilization patterns and identify potential areas for intervention.
7. Patient Outcomes:
Assess hospital outcomes such as diagnoses, readmission rates, and lengths of hospital stays to evaluate healthcare quality and patient outcomes comprehensively.
Through thorough extraction and analysis of these data elements, researchers and healthcare practitioners can derive valuable insights into patient care practices, healthcare utilization trends, and factors influencing healthcare outcomes, thereby informing evidence-based decision-making and enhancing patient care delivery.

Data Balancing

In classification tasks, imbalanced class distributions pose challenges to prediction modeling, often leading to skewed and less accurate results. Machine learning algorithms for classification typically begin with a balanced set of examples for each class under study to ensure fairness and accuracy in analysis. During preprocessing, efforts are made to handle incorrect data and missing values, which may include attributes like Patient-NumberAge, Diabetes Pedigree Function, Smoking, BMI, Insulin, Skin Thickness, Blood Pressure, Gender, and Glucose, ensuring completeness and consistency of the dataset. By addressing missing values and standardizing the dataset through scaling, a balanced representation of all features is achieved, enabling more reliable modeling outcomes.
The research in this domain has spurred significant advancements in resampling techniques aimed at mitigating class imbalances. Notably, techniques such as under sampling, which involves removing records from the majority class, and oversampling, which involves duplicating instances from the minority class, have emerged as effective strategies.

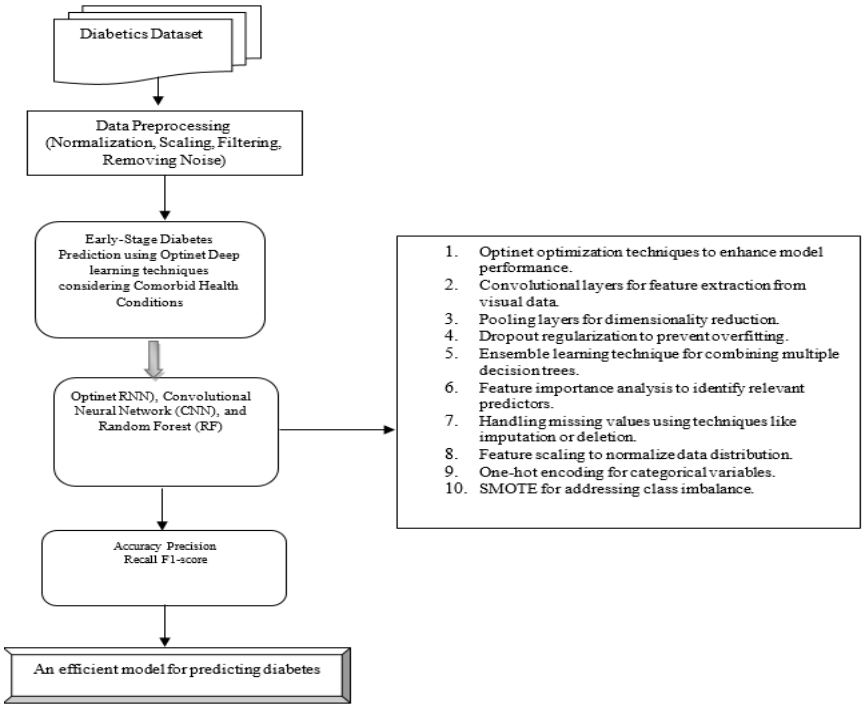

Figure 1 : Proposed Methodology

The above Figure 1 represents, The proposed methodology for early-stage diabetes prediction begins with data preprocessing, including normalization, scaling, filtering, and removal of irrelevant data. The processed data is analyzed using a combination of Optinet (RNN), CNN, and Random Forest models to improve accuracy. Techniques such as feature extraction using convolutional layers, dropout regularization, ensemble learning, and feature importance analysis are applied. Missing values are handled through imputation or deletion, and data distribution is normalized. Categorical variables are encoded using one-hot encoding, and SMOTE is used to address class imbalance. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics.

Feature Extraction Layers: This stage involves specialized modules dedicated to extracting relevant features from each input modality.

- **Numerical Data Processing:** Fully connected layers or recurrent layers (LSTMs) will process numerical data, identifying patterns and relationships within these variables. Normalization or standardization techniques will be applied to ensure that the data are appropriately scaled for optimal model performance.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 657-663

- **Categorical Data Processing:** Embedding layers will convert categorical variables into dense vector representations, enabling the model to learn meaningful relationships between these variables and the diabetes outcome.
- **Time-Series Data Processing:** LSTM layers are well-suited for capturing temporal dependencies within time-series data. These layers will learn patterns and trends in glucose levels or other relevant time-varying variables.
- **Image Data Processing:** A CNN module will analyze retinal images, extracting features that are indicative of diabetic retinopathy. Pre-trained CNN architectures (e.g., ResNet, Inception) can be used as a starting point, followed by fine-tuning on the specific dataset.

Fusion Layer: A critical component of Optinet is the fusion layer, which integrates the extracted features from different modalities. This layer could employ various techniques, including:

- **Concatenation:** Simply concatenating the feature vectors from different modalities.
- **Attention Mechanisms:** Using attention mechanisms to weigh the importance of different modalities in the prediction process. This allows the model to dynamically focus on the most relevant features for a given patient.
- **Multimodal Fusion Networks:** Employing more sophisticated multimodal fusion networks that learn optimal ways to combine features from different modalities.

Classification Layer: The final layer of Optinet is a fully connected layer with a sigmoid activation function. This layer outputs the probability of a patient having diabetes.

The specific configuration of Optinet, including the number of layers, activation functions, and hyperparameters, will be determined through rigorous experimentation and optimization using techniques like grid search, random search, or Bayesian optimization.

Cross Validation

Cross-validation in Deep Learning involves evaluating methods using a limited data sample. The parameter 'k' dictates the number of data groups generated from a given dataset and governs the evaluation process. This technique is commonly referred to as K-fold cross-validation.

K-Fold Cross Validation entails partitioning the dataset into K-folds, typically with values ranging between 5 and 10, depending on the dataset size. The model is then trained on K-1 folds and validated on the remaining fold (K-1), iteratively assessing the model's performance across different subsets of the data.

Table 1 : Performance Measures of Proposed Techniques

Performance metric	Formula
Recall	TP/(TP+FN)
Precision	TP/(TP+FP)
Accuracy	(TP+TN)/(TP+TN+FP+FN)

Table 1 displays performance metrics, including Recall, Precision, and Accuracy, with their respective formulas. Recall measures the classifiers' ability to correctly identify positive instances, Precision evaluates their capability to avoid misclassifying negative instances, and Accuracy assesses overall correctness in predicting early stage diabetes. These metrics offer insights into the effectiveness of the proposed techniques and guide further improvements in predictive modeling.

IV. Evaluation Metrics

Model performance was assessed using:

- **Accuracy**: Accuracy measures the proportion of correctly classified instances (both positive and negative) to the total number of instances, providing an overall view of the model's correctness. The formula for calculating accuracy is as follows:

$$Accuracy = (True\ Positives + True\ Negatives) / Total\ Instances \quad (1)$$

- **Precision:** Precision evaluates the proportion of correctly predicted positive instances out of all instances predicted as positive. It reflects the model's ability to minimize false positives. The formula for calculating precision is as follows:

$$Precision = True\ Positives / (True\ Positives + False\ Positives) \quad (2)$$

- **Recall:** Recall, also called sensitivity or true positive rate, measures the proportion of actual positive instances that the model correctly identifies, focusing on minimizing false negatives. The formula for calculating recall is as follows:

$$Recall = True\ Positives / (True\ Positives + False\ Negatives) \quad (3)$$

- **F1-Score:** The F1-score is the harmonic mean of precision and recall, providing a single metric to balance the trade-off between them, especially in imbalanced datasets. The formula for calculating F1-Score is as follows:

$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

- **ROC-AUC:** The ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) evaluates a model's ability to distinguish between classes. It measures the area under the ROC curve, where the curve plots the true positive rate (recall) against the false positive rate at various thresholds. The formula for calculating ROC-AUC is as follows:

$$AUC = \int_0^1 TPR\ d(FPR) \quad (5)$$

TPR (True Positive Rate) = True Positives / (True Positives + False Negatives)

FPR (False Positive Rate) = False Positives / (False Positives + True Negatives)

Classification of Algorithms

Optinet Deep Learning: Optinet Deep Learning stands out as a specialized variant within the neural network architecture, meticulously crafted to optimize the performance of these networks. Tailored with advanced optimization techniques like Optinet, this algorithm refines the training process, elevating model accuracy significantly. Its forte lies in tackling intricate datasets, making it an ideal choice for tasks demanding intricate feature extraction, particularly in domains like healthcare.

Random Forest: Random Forest emerges as a robust ensemble learning algorithm renowned for its prowess in handling high-dimensional datasets with intricate interactions. By constructing multiple decision trees during training, it offers a comprehensive approach to classification tasks, ensuring robustness and accuracy. Its resilience to overfitting renders it particularly suitable for applications such as early stage diabetes prediction, where precise classification is paramount.

Convolutional Neural Networks (CNNs): CNNs represent a cornerstone in deep learning, especially in tasks involving visual analysis. Proficient in discerning spatial patterns and hierarchies of features through convolutional filters, they excel in extracting meaningful insights from visual data. In the realm of early stage diabetes prediction, CNNs can be leveraged for feature extraction from medical images or pertinent visual data, facilitating the identification of crucial patterns indicative of diabetes onset.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 657-663

Average Performance Calculation

The average performance score provides a single metric to compare overall performance across classifiers:

Avg_Performance = (Precision + Recall + F1Score+Accuracy)/4

These algorithms offer complementary strengths and can be integrated to leverage their respective advantages in the development of predictive models for early-stage diabetes prediction.

Table 2: Various Deep Learning Algorithms performance comparison

Classifier	Precision	Recall	F1 score	Accuracy (%)
Optinet(RNN)	0.96	0.96	0.96	96%
CNN	0.91	0.91	0.91	91%
RF	0.89	0.89	0.89	89%

Performance Metrics

Each classifier is evaluated using:

- **Precision**: Proportion of true positive predictions among all positive predictions.
- **Recall**: Proportion of true positive predictions among actual positives.
- **F1 Score**: Harmonic mean of Precision and Recall.
- **Accuracy**: Proportion of correct predictions overall.

V. Results and Discussion

The results reveal commendable performance across all classifiers in predicting early stage diabetes, with Optinet (RNN) achieving the highest precision, recall, F1 score, and accuracy, each at 96%. This highlights the effectiveness of Optinet, a neural network variant, in enhancing model accuracy and performance. Its adeptness in handling intricate data and employing advanced optimization techniques contributes to its superior ability in accurately identifying early signs of diabetes onset. Close behind is the Convolutional Neural Network (CNN) with precision, recall, and F1 score all at 91%, showcasing its proficiency in analyzing visual data such as medical images for diabetes-relevant feature extraction. Random Forest (RF) also demonstrates respectable performance with precision, recall, and F1 score at 89%, indicating its effectiveness in managing high-dimensional data and complex interactions, albeit slightly lower than other classifiers. Despite minor performance metric variations, all classifiers demonstrate high accuracy, highlighting their potential in facilitating early diagnosis and intervention in diabetes management.

- Optinet (RNN) outperforms CNN across all metrics, achieving a higher score in Precision, Recall, F1 Score, and Accuracy.
- The performance gap is 0.05 (5 percentage points) in all metrics.
- Optinet (RNN) is more reliable for tasks requiring high precision and recall, which is critical for applications like medical diagnosis or fraud detection.
- CNN may still be suitable for tasks where a slightly lower accuracy is acceptable, possibly offset by other advantages (e.g., faster training or interpretability).

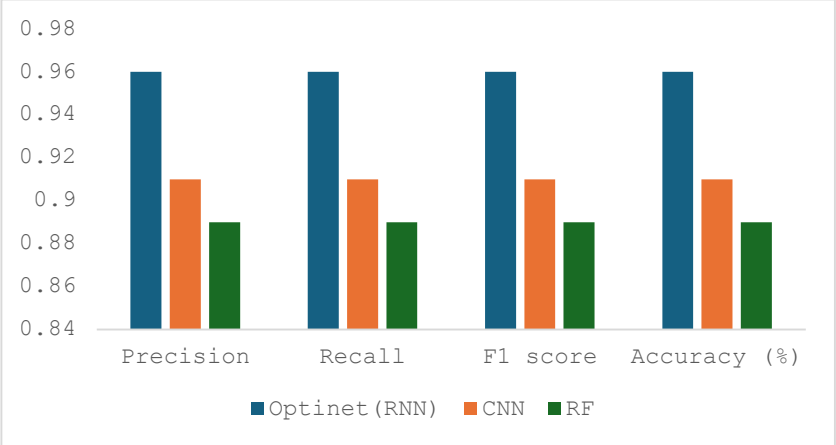

Figure 2: Performance comparison among ML Techniques

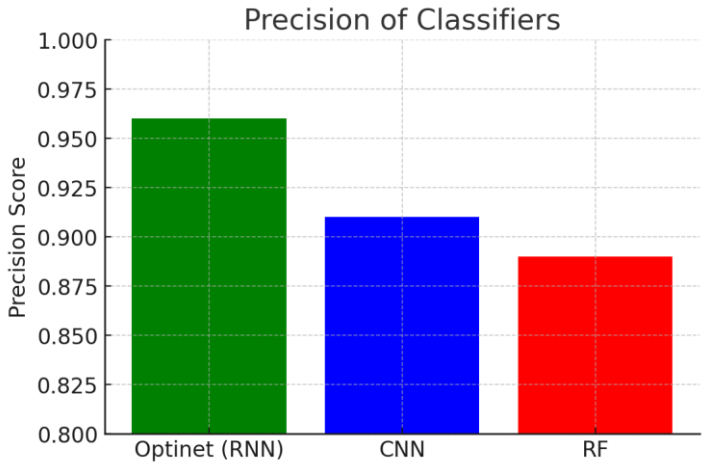

Figure 3 : Precision Classifiers

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 657-663

The figure 3 presents the Precision scores of the three classifiers: Optinet (RNN), CNN, and RF. Precision measures the model's ability to correctly identify positive cases among all predicted positive cases. Optinet (RNN) achieves the highest precision (0.96), indicating that it makes fewer false positive predictions compared to CNN (0.91) and RF (0.89). This demonstrates the effectiveness of RNN in capturing temporal dependencies within clinical data, leading to more precise diabetes predictions.

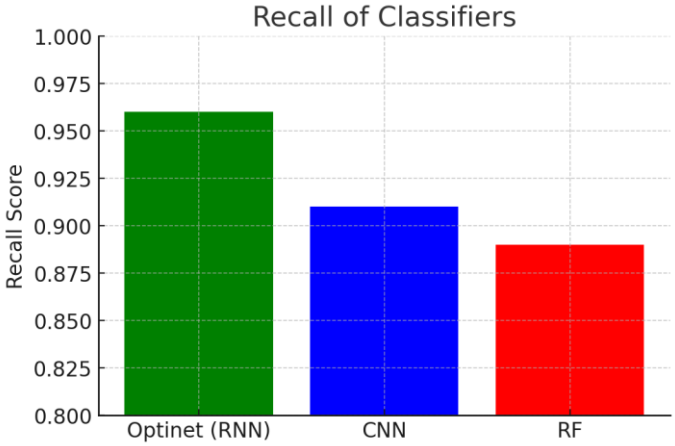

Figure 4: Recall Score

The Figure 4 illustrates the Recall scores, which measure how well each model identifies actual positive cases. A higher recall indicates that fewer true positives are missed. Again, Optinet (RNN) achieves the highest recall (0.96), followed by CNN (0.91) and RF (0.89). The superior recall of RNN is attributed to its ability to analyze sequential health records over time, improving the detection of patients at risk of diabetes.

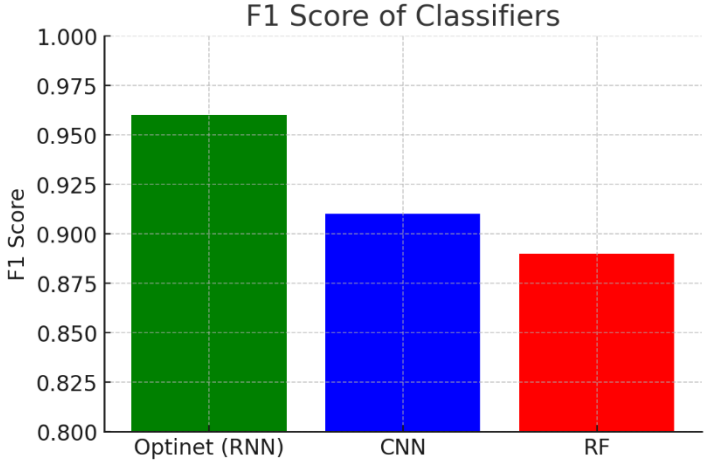

Figure 5: F1 Score

The Figure 5 showcases the F1 score, which is the harmonic mean of Precision and Recall. A high F1 score suggests a well-balanced model that minimizes both false positives and false negatives. Optinet (RNN) maintains the highest F1 score (0.96), reinforcing its robustness in diabetes prediction. CNN and RF exhibit lower F1 scores (0.91 and 0.89, respectively), suggesting that while they perform well, they are not as balanced as the RNN model in handling both false positives and false negatives.

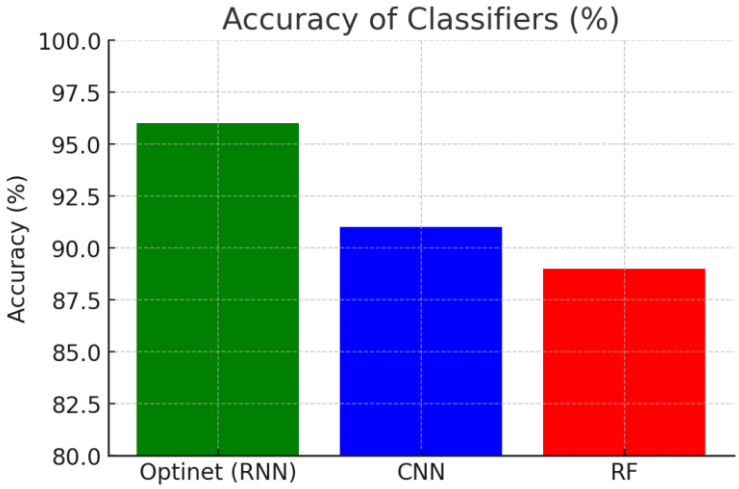

Figure 6: Accuracy

The figure 6 represents the Accuracy (%) of each classifier, indicating the overall proportion of correct predictions. Optinet (RNN) achieves the highest accuracy (96%), followed by CNN (91%) and RF (89%). The higher accuracy of Optinet (RNN) highlights the advantage of leveraging optimized recurrent architectures for long-term patient data analysis, ensuring more reliable diabetes risk assessment.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 657-663

From the performance metrics, it is evident that Optinet (RNN) outperforms CNN and RF in all aspects. The superior Precision, Recall, F1 score, and Accuracy demonstrate the effectiveness of RNNs in analyzing temporal patterns in clinical data. While CNNs are effective in feature extraction from structured data and Random Forests provide good interpretability, deep learning-based temporal models (RNNs) prove to be the most effective for early diabetes prediction.

The outcomes affirm the efficacy of deep learning algorithms, notably Optinet and CNN, in predicting early stage diabetes, thereby offering promising avenues for leveraging advanced computational techniques in healthcare. The high precision, recall, and F1 scores exhibited by all classifiers underscore their capacity to accurately identify individuals at risk of diabetes onset, enabling timely intervention and tailored healthcare strategies. These findings underscore the importance of harnessing sophisticated machine learning techniques, specifically tailored to the unique intricacies of healthcare data, in tackling complex medical conditions like diabetes.

VI.Conclusion

This research work examined how well the Random Forest (RF), Convolutional Neural Network (CNN), and Optinet (RNN) classifiers predict diabetes in its early stages. Optinet was found to have the highest precision, recall, F1 score, and accuracy, with scores of 96%. CNN and RF came in second and third, respectively, and third. These findings highlight the ability of deep learning algorithms, especially CNN and Optinet, to precisely identify people who are at risk of developing diabetes, allowing for prompt intervention and individualized treatment plans.

The approach used included sophisticated machine learning algorithms for feature extraction and classification in addition to stringent data pretreatment methods for managing missing values and class imbalance. The rigorous model assessment and validation provided by the application of cross-validation techniques further strengthened the dependability of the study's findings. All things considered, these results demonstrate the potential benefits of applying advanced machine learning methods specifically designed for healthcare data. The classifiers' potential to improve patient outcomes and healthcare delivery is indicated by their high accuracy, recall, and F1 scores. In order to improve the use of machine learning in healthcare and enhance predictive modeling skills for early illness identification and management, further research and development in this area will be essential.

References

1. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
2. Christodoulou, E., Ma, J., Collins, G. S., & Steyerberg, E. W. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of Clinical Epidemiology, 110, 12-22.
3. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.
4. Gujral, U. P., Mohan, V., Pradeepa, R., Deepa, M., Anjana, R. M., & Narayan, K. M. V. (2017). Ethnic differences in the prevalence of diabetes in underweight and normal weight individuals: The CARRS and NHANES studies. Diabetes Research and Clinical Practice, 123, 106-112.
5. Kumar, A., Jain, V., Kumar, A., & Kaur, R. (2019). Early Prediction of Diabetes Using Machine Learning Techniques: A Comprehensive Review. In 2019 IEEE 10th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT) (pp. 1-6). IEEE.
6. Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzel, R. (2015). Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:1511.03677.
7. Martínez-Castillo, J., Cárdenas-Ovando, E., Valdés-Pérezgasga, F., & Rizo, D. O. (2018). Classification of Diabetes Mellitus Using Machine Learning Techniques. Journal of Computer Science and Technology, 19(3), 133-147.
8. Nagarajan, R., Suthakar, J., & Manogaran, G. (2017). An enhanced predictive model for diabetes using effective features. Health Information Science and Systems, 5(1), 4.
9. Natarajan, N., Pu, J., Liu, H., Alizadeh, A. A., & Wei, W. Q. (2017). Gene expression biomarkers that discriminate early stage lung adenocarcinoma from normal lung and lung adenocarcinoma from squamous cell carcinoma. American Journal of Respiratory and Critical Care Medicine, 195(4), 492-503.
10. Oh, S. L., Ng, E. Y. K., & Acharya, U. R. (2020). A deep learning approach for automatic detection of diabetes retinopathy. Applied Sciences, 10(4), 1276.
11. Razavian, A. S., Marcus, J., & Sontag, D. (2016). Multi-task prediction of disease onsets from longitudinal laboratory tests. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1105-1114).
12. Razavian, A. S., Sontag, D., & Marcus, J. (2015). Time-series abnormality detection algorithms for clinical data. Journal of Biomedical Informatics, 54, 43-55.
13. Shaw, J. E., Sicree, R. A., & Zimmet, P. Z. (2017). Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Research and Clinical Practice, 87(1), 4-14.
14. Singh, A., & Shukla, A. (2020). Early Prediction of Diabetes Mellitus using Machine Learning Algorithms: A Review. In 2020 International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) (pp. 261-267). IEEE.
15. Zhang, Y., Wu, Q., & Zhu, X. (2019). An Automatic Prediction Model of Diabetes Using Machine Learning Algorithms. In 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 1494-1499). IEEE.
16. Zhao, M., Ren, Z., Li, H., Zhang, J., & Han, J. (2020). Time-aware deep adversarial network for early diabetic kidney disease prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17(2), 675-686.
17. Abdulnaser Rashid, Mohana Priya T, Abdalla Ibrahim Abdalla Musa, Suliman Mustafa Mohamed Abakar, Siti Sarah Maidin, Rajesh Kanna R, Mahalakshmi S B(2026), A Comprehensive Review of Artificial Intelligence Methods for Tumor Detection in Pancreatic Cancer Diagnosis. (2026). *MSW Management Journal*, *36*(1s), 394-396.