Available online at https://icsejournal.com/index.php/JCSE
Journal of Computer Science an Engineering (JCSE)
e-ISSN 2721-0251
Vol. 4, No. 2, August 2023, pp. 61-77

An Integrated Approach for Diabetes Detection Using Fisher Score Feature Selection and Capsule Network

Mohammad Sadman Tahsin[1*], Musaddiq Al Karim[2], Minhaz Uddin Ahmed[3], Faiza Tafannum[4], and Neda Firoz[5].

[1,3] BRAC University, 66, Dhaka 1212, Bangladesh
[2] American International University – Bangladesh (AIUB), 408/1, Dhaka 1229, Bangladesh.
[4] University of Texas at Arlington, 701 S Nedderman Dr. Arlington, TX 76019, United Stated of America.
[5] Tomsk State University, Tomsk, Lenin Ave, 36, Tomsk, Tomskaya oblast', Russia
[1]sadman.tahsin3@gmail.com*; [2]musaddiqmk19@gmail.com; [3]minhazmashfi326@gmail.com; [4]fxt8308@mavs.uta.edu; [5]nedafiroz1910@gmail.com.
* corresponding author

ARTICLE INFO

Article History:
Received September 7, 2023
Revised September 23, 2023
Accepted October 1, 2023

Keywords:
Diabetes Detection
Capsule Network
Machine Learning
Fisher Score
Classification

Correspondence:
E-mail: sadman.tahsin3@gmail.com

ABSTRACT

This paper investigates how the Fisher score feature selection approach can be used with capsule networks for diabetes detection. It also evaluates how well this algorithm works based on a number of evaluation parameters. The selected features using Fisher score method was then employed to train a capsule network model. Accuracy (94%), precision (94%0, recall (94%), F1 score (94%), and other performance evaluation metrics were thoroughly analyzed to determine the algorithm's efficacy. The results demonstrated that the combination of Fisher score feature selection and capsule networks yielded promising performance in diabetes detection. The selected features effectively captured the relevant information necessary for accurate classification The capsule network model was very accurate, which shows that it could be a good tool for diagnosing diabetes. Also, the accuracy and recall values showed that the algorithm could correctly place both positive and negative cases of diabetes, minimizing the risk of misdiagnosis. By merging the Fisher score feature selection approach with capsule networks, this research study contributes to advancing diabetes detection.

1. Introduction

Insufficient insulin production by the pancreas or resistance to the effects of insulin both contribute to the development of diabetes, a prevalent and debilitating chronic disease. As a result, there is an elevation of glucose concentration in the bloodstream, known as hyperglycemia [1]. According to [2] the global prevalence of diabetes would rise from 382 million in 2013 to 592 million by 2035. Heart disease, blindness, kidney failure, and amputations are all potential outcomes of this disorder, making it complex enough to warrant regular monitoring and treatment. In order to effectively manage diabetes, early detection, and intervention are essential, but diagnosing the disease can be difficult because there are frequently no symptoms in the early stages. Diabetes translates roughly as excessively sugary urine. The three kinds of diabetes are Type 1, Type 2, and Gestational Diabetes. Lack of physical labor, obesity, and lifestyle modifications are the primary causes. Diabetes symptoms include polydipsia and polyuria, among others [3]. The fundamental defect is a relative or absolute absence of insulin, the pancreatic hormone that promotes the transfer of glucose into tissues for glycogen or fat storage. The infiltration of insulin-producing cells by the immune system is the underlying cause of type 1 diabetes, which is treated with insulin therapy. Type 2 diabetes is caused by a relative insulin shortage typically coupled by tissue

http://dx.doi.org/10.36596/jcse.v4i1.708

insulin resistance. It is a condition that progresses slowly, so although diet and lifestyle adjustments are often effective initially, oral medication and insulin may be required in the long run. Long-term vascular complications affecting both small and large arteries are associated with diabetes [4].

One of the foundations of AI is machine learning (ML), which builds on prior knowledge to improve the performance of AI systems [5]. Machine learning attempts to address the question of how to make computers that get better on their own as they use them [6]. For a variety of perceptual tasks, machine learning systems can learn to map inputs to outputs [7]. It includes training a model on a collection of labeled data and then using that model to generate predictions on additional, unlabeled data. ML's two basic components are Supervised Machine Learning and Unsupervised Machine Learning [8]. Supervised Learning is the most prevalent kind of machine learning. During the training phase, it labels each of the system's inputs with the value it wants to get out of them. Another popular type of machine learning is "Unsupervised Learning", which makes conclusions without the idea of labels [9]. The majority of researchers employ linear regression, support vector machines, random forests, and naïve Bayes, as well as other comparable techniques. [10]–[14]. Cluster analysis, K-means, the Apriori algorithm, etc., are some of the most popular ways to learn without being watched [15]–[17]. Some of deep learning algorithms were also applied for diabetes detection [18], [19]. Several challenges in healthcare, including the detection of diabetes, have been overcome via machine learning. There are multiple perks to using machine learning algorithms for diabetes detection. First, it can uncover patterns in large, complex datasets that are not readily apparent to humans. Second, it can provide timely and accurate diagnoses, allowing for early intervention and improved outcomes. Thirdly, it can reduce the need for time-consuming and inconvenient diagnostic tests that are intrusive or costly.

1.1. Related works

Diabetes disease draws considerable interest in the machine learning community. Various machine learning methods, such as DT, Random Forest, LR, Discriminant Analysis, SVM, kNN, ensemble learners, etc., are used for early stage diabetes detection [20]–[25]. Various methods, such as 10-fold cross-validation [26], average classification accuracy [27], and and so on, were used to evaluate the effectiveness of the results.

The Pima Indians Diabetes Data Set, which contains data on patients with and without diabetes, was used in [28]. Numerical discretization, missing value management, and attribute selection and identification are all part of the initial stage of data preparation. The creation of a diabetes prediction model utilizing the decision tree algorithm is the second stage. To create a decision tree-based diabetes prediction model, Rapid-I's RapidMiner is also applied to the Pima Indians dataset's diabetes data. [29]. Using feature selection techniques such as recursive feature elimination [30], the genetic algorithm (GA) [31], and the Boruta package [32], the performance of decision tree classifiers was enhanced. After applying the feature selection technique, the PID dataset was utilized to evaluate the model's performance [33]. Random forest is one of the most recent and fruitful findings in decision tree learning research. It is widely employed in the medical field, especially for diagnosing diabetes [34]. Using a specific parameter for the Random Forest algorithm, from which a binary mask of exudate is derived after intensity thresholding, the sensitivity is 91.40 percent and the accuracy is 94.38% [35]. Using the Minimal Redundancy Maximal Relevance feature selection method to select features from the Pima Indians Diabetes Data and applying seven distinct types of performance metrics using a 10-fold cross-validation approach, a remarkable accuracy was attained through the Random Forest approach. [36].

http://dx.doi.org/10.36596/jcse.v4i1.708

Using the same dataset, Naive Bayes achieved 76.30% accuracy. The ROC curve verified the results [37]. Using a different dataset, the classical Naive Bayes model yielded an accuracy rate of 87.69%, which increased to 88.65% when the GA feature selection method was applied [38]. Another comparison study of [39] revealed that Naive Bayes had a 78.57% accuracy [40]. The Restricted Boltzman machine is a basic building block of the Deep Network which can be used as a classifier as well as feature extractor to obtain better results. Variational autoencoder has been used for data augmentation and Sparse autoencoder for feature augmentation on the Pima Indians Diabetes Database. Convolutional Neural Network has been used for classification which yielded 92.31% accuracy on detecting diabetes [41]. Comparison of Gaussian Naïve Bayes, Linear Support Vector Machine and the Random Forest through the lens of predictive performance accuracy and Area Under Curve score demonstrated that LSVM showed best accuracy with 78.39% and Random Forest performed the least with an average accuracy of 72.72% [42]. Enhanced Support Vector Machine has been applied on a specific dataset and the output has been used as an input through Deep Neural Network to achieve a combined efficacy [43]. A comparison of the performance of two-class logistic regression, the two-class neural network, the two-class decision jungle, and the two-class boosted decision tree on a dataset containing 15,000 women aged 20 to 80 revealed the two-class boosted decision tree to be the most accurate model [44]. LR, RF, and a DNN containing the embeddings for the categorical features have been applied in a specific dataset. DNN with embeddings showed promising results with an F1 score of 1.0 on the test data [45]. Extreme gradient boosting has been applied as classifier using ECG signals as inputs to predict patients with type-2 diabetes or impending type-2 diabetes. The algorithm accurately predicted the classes with 97.1% precision, 96.2% recall, 96.8% accuracy, and 96.61% F1 score [46]. The main limitation of the proposed models is that the lack of analysis of performance evaluation metrics. This paper analyzed many performance evaluation matrices like accuracy, precision, recall, f1-score etc. in order to derive a better prediction model using which is absent in most of the works related to diabetes prediction as they have taken selective performance evaluation matrix for their predictions. The following are the paper's contributions:

1. Explore and analyze the efficacy of Fisher score feature selection combined with capsule networks for diabetes detection.
2. Evaluate the performance of the capsule network model trained on selected features, focusing on accuracy, precision, recall, and F1 score. Assess the method's ability to appropriately classify diabetes cases, reducing the risk of misdiagnosis.
3. Showcase the promising performance of the Fisher score feature selection and capsule networks approach in diabetes detection.
4. Highlight the potential of the proposed approach to enhance the accuracy and efficiency of diabetes detection, ultimately leading to personalized treatment options and improved patient outcomes.

2. Methods

2.1 Proposed System Architecture:

This paper utilizes a dataset which was collected from a hospital in Sylhet. Important feature has been extracted from the dataset. The capsule network has been developed using multiple layers. All the values have been converted to numerical values. The performance of the Capsule Network has been measured in terms of precision, F1-score, recall etc.

http://dx.doi.org/10.36596/jcse.v4i1.708

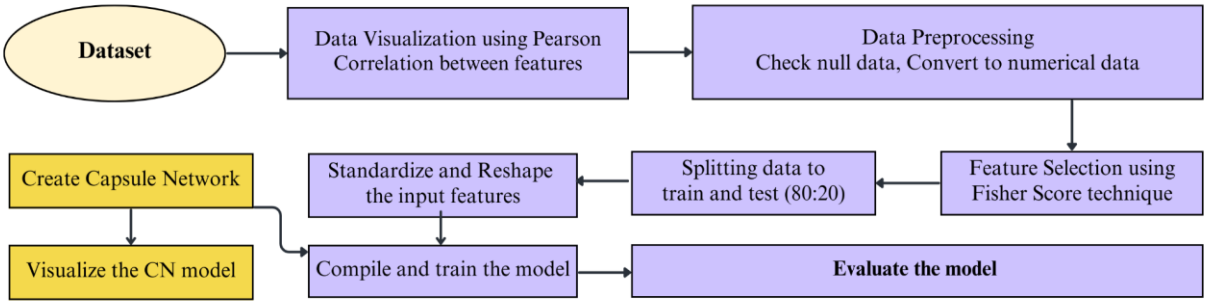

Figure 1. Architecture of the proposed system.

2.2 Dataset Description

As diabetes is a critical disease new data were collected utilizing existing benchmark datasets, 520 individuals' diabetes-related symptoms are included in this data set. It includes details about individuals, such as symptoms that may contribute to diabetes. This dataset was collected through a direct survey of individuals just diagnosed with diabetes or who are not yet diabetic but exhibit few or no symptoms. Sylhet Diabetes Hospital in Sylhet, Bangladesh, patients submitted the data [47]. The following information describes various aspects of a dataset related to diabetes, shown in figure 2.

1. **Age**: The data was collected from a diverse population spanning a wide age range, encompassing individuals aged between 20 and 90 years.
2. **Gender**: There are total of 520 individuals and among them, 328 are male and 192 are female.
3. **Polyuria**: Increased urination is a common diabetes symptom. In response to elevated blood glucose levels, the kidneys must work harder to filter and absorb extra glucose, resulting in an increase in urine production [48].
4. **Polydipsia**: Another usual sign of diabetes is having to drink a lot. This happens because the body is losing water because of increased urine production [49].
5. **Sudden weight loss**: Unintentional weight loss can be a symptom of diabetes that has not been diagnosed. This is a result of the body's inability to correctly utilize glucose for energy, causing it to rely on fat reserves instead. It is usually found on type 1 or juvenile diabetes mellitus [50].
6. **Weakness**: Due to the body's inability to utilize glucose as an efficient source of energy, diabetes may induce symptoms such as weakness and weariness [51].
7. **Polyphagia**: Increased appetite is another diabetes symptom. This occurs when the body's cells do not receive enough glucose, causing inadequate stimulation of the satiety center and causing the body to crave more nutrition [52].
8. **Genital thrush**: Diabetes may affect the immune system, increasing susceptibility to fungal infections such as thrush [53].
9. **Visual blurring**: When blood sugar levels are too high, the shape of the eye lens might alter, which can make it difficult to see [54].
10. **Itching**: Itchiness of the skin is a prevalent complaint among diabetics, which may be due to poor circulation or nerve injury [55].
11. **Irritability:** Mood swings and irritability are two signs that diabetics may experience as a direct result of fluctuations in their blood sugar levels [56].
12. **Delayed healing**: High blood sugar levels can damage blood vessels and neurons, resulting in inefficient circulation and slowed wound healing [57].
13. **Partial paresis**: The damage to the nerves that diabetes causes may sometimes lead to a weakening of the muscles or even partial paralysis [58].

http://dx.doi.org/10.36596/jcse.v4i1.708

14. **Muscle stiffness**: Because of the way that diabetes alters the way that the body processes glucose, it is possible for diabetics to have tight muscles [59].
15. **Alopecia**: Diabetes may cause poor circulation as well as hormonal abnormalities, both of which can lead to hair loss or thinning [60].
16. **Obesity**: Obese persons are more prone to acquire type 2 diabetes, which occurs when the body develops insulin resistance. Being obese is a risk factor [61].
17. **Class**: 'Diabetes Positive' and 'Diabetes Negative' are the two classes of this research work.

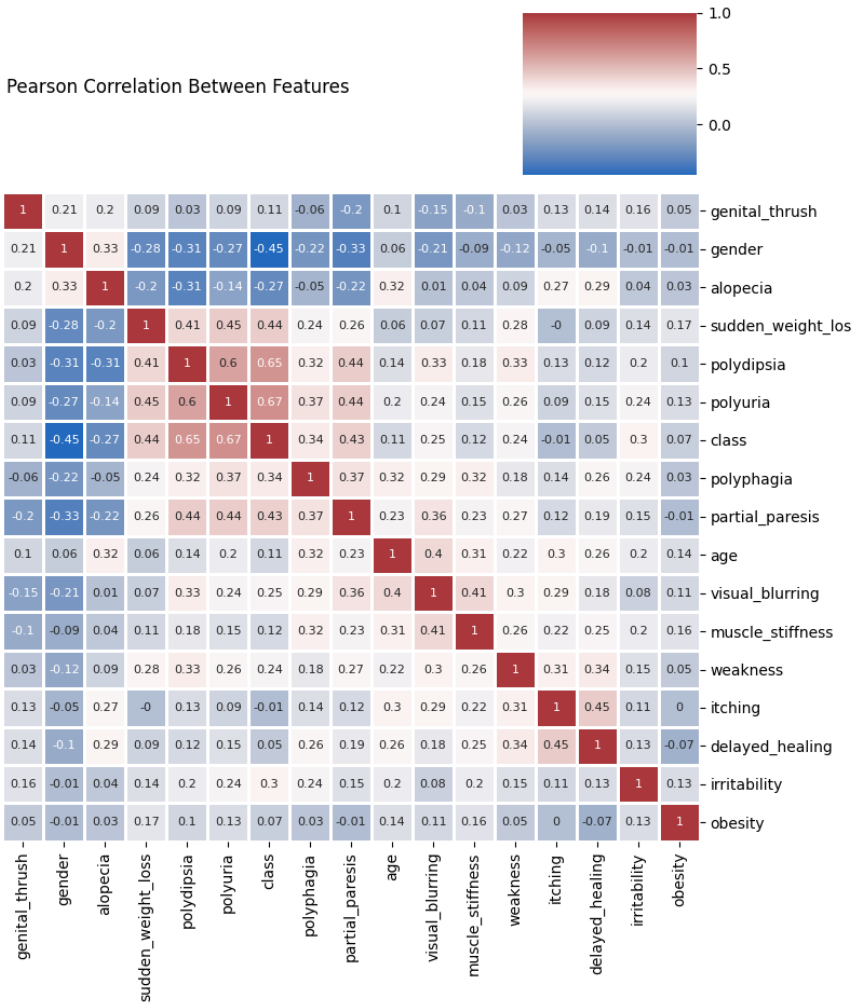

Figure 2. Pearson Correlation between Features.

2.3 Feature Selection

Fisher Score: Fisher Score is one of the typical supervised feature selection approaches that aims to identify the most efficient features that reduce within-class scatter and maximize between-class scatter [62]. This method is useful selecting features that are not only relevant but also the independent from each other. As Fisher sccore is a linear discriminant method, it's effective at capturing linear relationships between features and class labels. Each feature is given a score based on its ability to distinguish between classes. The higher the score, the greater the feature's classification relevance. Features with better scores are thought to be more useful for classifying and are chosen for the final model [63]. This method works well when the dataset size is small to medium. As the experimental dataset size is medium, this method

http://dx.doi.org/10.36596/jcse.v4i1.708

is effective for selecting key features. The method can boost the performance of the classification model and lower the risk of overfitting. By implementing Fisher score feature selection method, 11 attributes were selected.

2.4 Proposed model:

2.4.1 Capsule Network

Since the 2011 development of the concept, capsule networks have received significant interest. They perform better than CNN at acquiring viewpoint-invariant representations and modeling part-to-whole relationships between entities [64]. CNN commonly uses pooling layers for the purpose of down sampling and abstracting features. However, it is important to note that this process can result in a reduction of spatial information. Capsule Networks (CapsNets) have the potential to decrease the necessity of pooling layers, which might be beneficial in situations where the inclusion of intricate features is crucial for accurate diagnosis. In certain cases, Capsule Networks (CapsNets) sometimes require a smaller number of parameters compared to Convolutional Neural Networks (CNNs) in order to attain comparable performance levels. This characteristic can result in the development of computationally efficient models, which is particularly significant in scenarios that demand real-time diagnosis or operate within limited resource contexts. CNN neglect the critical spatial link between simple and complex objects. Instead of the initial translational invariance, CapsNets now gain equivariance using a new design that mimics the human visual system, which allows them to require less data to achieve a more thorough generalization across various angles [65]. It has been claimed that Capsule Networks exhibit a reduced vulnerability to overfitting, a particularly crucial aspect when working with limited medical data. As the dataset size is small, this model is well suited for the experiment. A capsule is a set of neurons whose output predicts specific aspects of the same item. Each layer of a capsule neural network consists of several capsules [66]. Capsule Neural Network, also known as CapsNet, is a form of ANN used in machine learning systems to represent hierarchical relationships [67], [68].

The model uses a shape input_shape as input and consists of a series of convolutional and dense layers. The convolutional layers (conv1 and primarycaps) extract features from the input data, whereas the dense layers (digitcaps and output) create the core of the Capsule Network. The digitcaps layer represents the output capsules that convey the presence and characteristics of different classes. Calculating the output involves computing the square root of the sum of squared activations along the capsule dimensions.

2.4.2. Structure of the model

Proposed Capsule Network model, shown in figure 3.

http://dx.doi.org/10.36596/jcse.v4i1.708

 e-ISSN 2721-0251

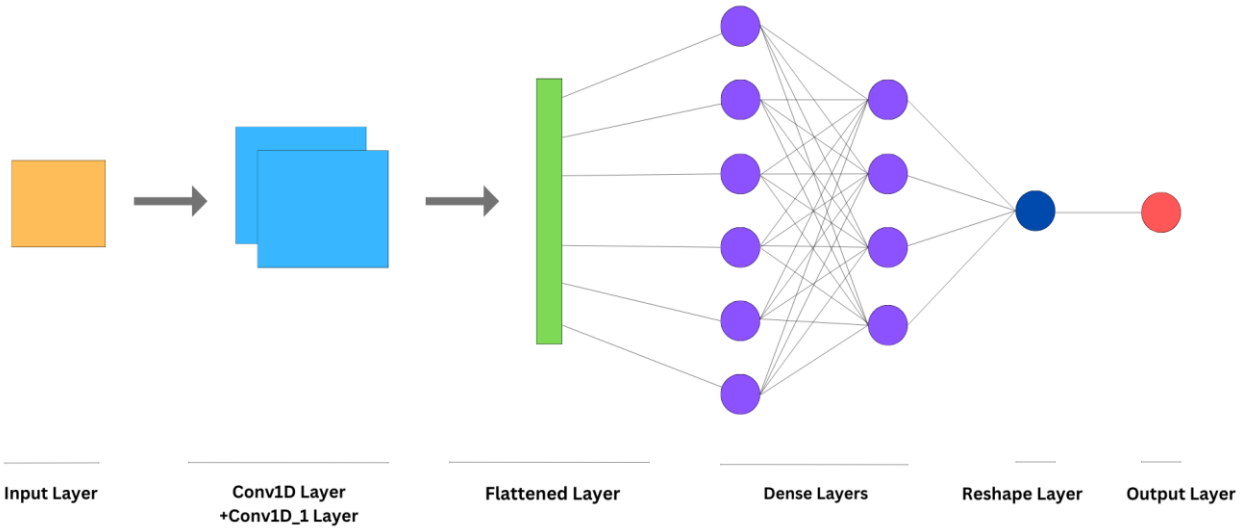

Figure 3. Proposed Capsule Network model.

Breakdown of each layer in the capsule network model, including the layer type, output shape, and number of parameters:

1. "input_1": InputLayer with an output shape of '(None, 11, 1)', indicating that the input data has a shape of '(batch_size, 11, 1)'. The None dimension represents the batch size, which can vary during training.
2. "conv1d": Conv1D layer with an output shape of '(None, 7, 256)'. This layer performs convolutional operations on the input, resulting in a feature map with a width of 7 and 256 channels. It has 1,536 parameters.
3. "conv1d_1": Conv1D layer with an output shape of '(None, 3, 256)'. This layer performs another convolutional operation on the previous layer's output, reducing the width further to 3. It has 327,936 parameters.
4. "flatten": Layer that flattens the output of the preceding layer into a 1D form vector. '(None, 768)'.
5. "dense": Dense layer with an output shape of '(None, 256)'. This layer applies a fully connected operation with 256 units to the flattened input. It has 196,864 parameters.
6. "dense_1": Dense layer with an output shape of '(None, 16)'. This layer applies another fully connected operation, reducing the dimensionality to 16. It has 4,112 parameters.
7. "reshape": Reshape layer with an output shape of '(None, 1, 16)'. This layer reshapes the previous layer's output to have a shape of '(None, 1, 16)'.
8. "output": Lambda layer that applies a lambda function to compute the final output. It produces a form that is '(None, 1)'.

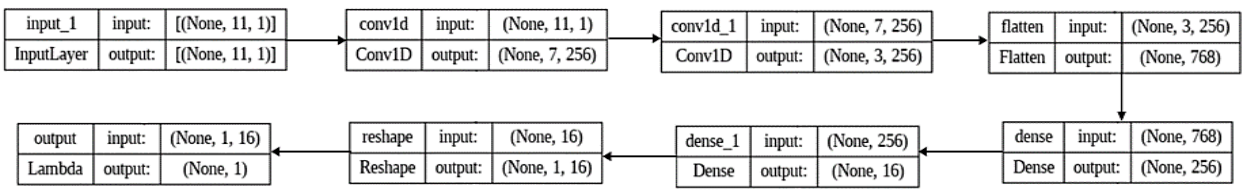

Figure 4. Structure of the layers.

http://dx.doi.org/10.36596/jcse.v4i1.708

2.4.3. Classification with Capsule Network

A large number of training samples are used to train the network so that it can categorize co-channel signals efficiently. The output of the capsule network is an eight-dimensional vector denoted by v = [vi], where i ∈ [1, 8] and vi ∈ (0, 1). The threshold decision function f (), represented as determines the final classification result.

$$f(x) = \begin{cases} 1 \; if \; x \; \in \; [\mathrm{T}, 1] \\ 0 \; if \; x \; \in \; [0, \mathrm{T}] \end{cases} \quad (1)$$

Where, T is a manually set parameter for decision-making.

2.5. Performance Evaluation Parameter

TP, TN, FP, and FN are used to calculate the accuracy of classifiers [69]. Precision and recall contribute as statistical validation metrics [70]. Precision indicates the positive predictive value of the classifier. While recall is the percentage of positive test results, it is the ratio of precisely predicted positive observations to all accurately predicted positive observations [71].

$$Precision = (TP/(TP + FP)) \times 100 \quad (2)$$
$$Recall \; (Sensitivity) = (TP(TP + FN) \times 100) \quad (3)$$
$$Accuracy = ((TP + TN)/(TP + TN + FP + FN)) \times 100 \quad (4)$$

The ROC is a graphical representation of the performance of a classifier as thresholds increase [72]. Between each specificity and sensitivity value is the ROC curve [73]. The significance of ROC is proportional to the number of tested thresholds. F-measure is an additional statistical performance indicator [74].

$$F - measure = ((2 \times TP)/(2 \times TP + FP + FN)) \quad (5)$$

The numbers in a confusion matrix represent the disparity between expected and actual values [75]. The "TN" output indicates the number of samples that were correctly identified as negative. Similarly, "TP" indicates the number of correctly identified positive events. As "FP" implies, the number of erroneously positive occurrences equals the number of falsely negative occurrences; similarly, "FN" suggests that the number of falsely negative occurrences equals the number of falsely positive occurrences. In the context of classification, precision is an essential performance metric.

The Cohen's Kappa coefficient, also known as the Kappa score, is a statistical indicator of inter-rater reliability for categorical or nominal data [76]. It is typically used to assess the level of agreement between anticipated and real labels in machine learning and classification projects.

$$Kappa \; = \; (Po \; - \; Pe) \; / \; (1 \; - \; Pe) \quad (6)$$

Where, (i) Po is the observed agreement, which is the proportion of cases where raters or classifiers agreed, (ii) Pe denotes the proportion of cases in which agreement is expected by chance.

The Kappa score has a range of -1 to 1. A value of 0 denotes agreement by chance, a value of 1 denotes perfect agreement, and a value of -1 denotes disagreement that exceeds chance. When comparing and evaluating classification models, especially when dealing with imbalanced datasets or taking the chance agreement into account, the Kappa score is widely used [77].

http://dx.doi.org/10.36596/jcse.v4i1.708

MSE, RMSE, and MAE have frequently employed performance metrics used to evaluate the accuracy and error of prediction models. Here is a breakdown of each metric:

a. Mean Squared Error (MSE): The average squared variation between expected and observed values is what the MSE calculates. The disparities between the expected and actual values are called squared residuals, and this calculation finds their mean. Due to the squaring procedure, MSE penalizes greater mistakes more severely [78].
b. Root Mean Squared Error (RMSE): The RMSE is an easier-to-understand metric than the MSE due to the fact that it shares a unit with the objective variable (the square root of MSE). The Root Mean Squared Error (RMSE) is used to measure how large the model's errors normally are [79].
c. Mean Absolute Error (MAE): The MAE is a statistical measure of how far predictions deviate from actual results. Due to the non-squaring of the errors, MAE is less affected by extreme values. The MAE provides a more concise explanation of the typical size of errors [80].

3. Result analysis

The effectiveness of the model for diabetes identification is thoroughly examined in this section. The evaluation measures used in this study include MSE, RMSE, MAE, MSE, Recall, F1-Score, Overall Accuracy, and Kappa Score. Also given are the values for Accuracy, Precision, and Recall for each particular class. These measurements were used to evaluate the model's effectiveness in detecting diabetes.

3.1. Performance Metrics

Table 1. Summarizes the Performance Metrics Obtained from this Model.

Evaluation Metric	Result
Precision	0.94
Recall	0.94
F1-Score	0.94
Class wise Precision	(Class 0: 0.91, Class 1: 0.96)
Class wise Recall	(Class 0: 0.91, Class 1: 0.96)
F1-Score (Class 0)	0.91
F1-Score (Class 1)	0.96
Validation Accuracy	0.94
Validation Loss	0.39
Overall Accuracy	0.94
Kappa Score	0.87
MSE	0.058
RMSE	0.24
MAE	0.058

Performance Evaluation of the Classification Model: An evaluation of a classification model's performance is frequently done using the precision, recall, and F1-Score metrics. A balanced performance in accurately detecting both positive and negative cases is indicated by the acquired precision (0.94) and recall (0.94) values. The robustness of the model is shown by the F1-Score's (0.94) strong harmonic mean between precision and recall. Impressive results are also seen when class-level precision, recall, and F1-Score are examined for each target class. In comparison to Class 0 (positive cases), Class 1 (diabetes) displays somewhat higher precision (0.96) and F1-Score (0.96). This shows the model's effective classification between two classes and its successful identification of people with diabetes and normal instances. **Overall Accuracy:** The model's validation accuracy and overall accuracy of 0.94 indicate that its predictions were accurate to a significant degree. The model's validation loss is

http://dx.doi.org/10.36596/jcse.v4i1.708

0.39. This statistic offers a comprehensive analysis of the model's performance across all classes and acts as a trustworthy gauge of its usefulness in practical situations. **Agreement and Error Analysis:** The Kappa score was produced to assess the degree of agreement between the model's predictions and the ground truth labels. A Kappa value of 0.87 was attained, suggesting significant agreement that was not the result of random chance. This emphasizes how accurate and consistent the model's predictions are. Additionally, a review of error measurements, such as MSE, RMSE, and MAE, was done. The model's correctness is further supported by the low MSE (0.058), RMSE (0.24), and MAE (0.058) values that were found. These measures show little variation between the expected and actual values, demonstrating the accuracy of the model.

3.2. Confusion Matrix

With 30 occurrences accurately classified as Class 0 and 68 cases correctly classified as Class 1, the diagonal of the matrix shows the correctly anticipated classes. However, three instances from Class 0 and three examples from Class 1 were incorrectly categorized as being in the same class. With a few misclassifications, the model performs well overall in properly categorizing examples from both Classes 0 and 1. It shows a higher accuracy in identifying Class 1 instances, as evident from the larger number of correct predictions for this class. This interpretation provides insights into the model's performance in differentiating between the two classes and highlights areas where misclassifications occurred. The confusion matrix for the classification results of the model is as shown in figure 5.

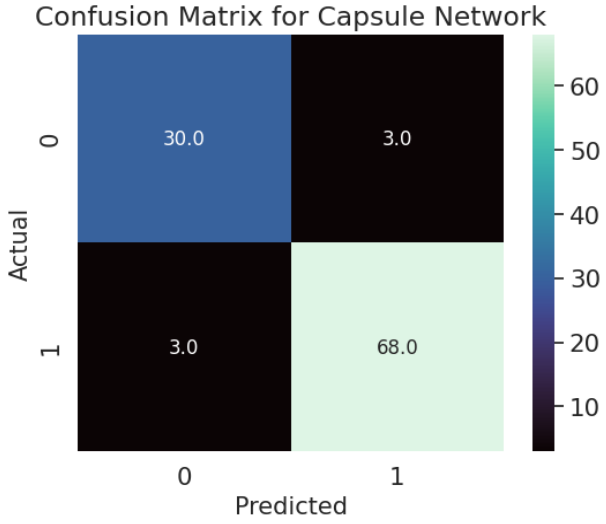

Figure 5. Confusion Matrix.

3.3. Receiver Operating Characteristic (ROC)

The ROC curve resulted in an AUC value of 0.93, indicating the model's robust ability to differentiate between positive and negative instances in the diabetic detection task. The AUC value, which falls within the range of 0 to 1, provides a comprehensive evaluation of the model's performance across diverse classification thresholds. By achieving an AUC of 0.93, the model demonstrates a notable level of accuracy and effectiveness in accurately prioritizing and categorizing instances. A higher AUC value implies an enhanced capability to correctly classify instances across different thresholds.

http://dx.doi.org/10.36596/jcse.v4i1.708

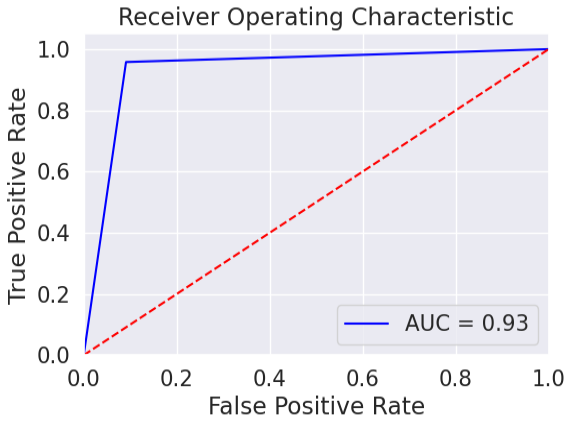

Figure 6. ROC Curve.

4. Discussion

The precise and well-timed detection of diabetes is of supreme importance for effective disease management and treatment. Conventional methods of diabetes detection often rely on physical examination and individual interpretation, which can be susceptible to humanoid error and variability. Amidst recent developments, machine learning techniques have displayed potential in improving the accuracy and efficiency of diagnosing diseases as critical as diabetes. This work focuses on the deployment of the Fisher score feature selection approach in combination with capsule networks for diabetes detection. The Fisher score method, a widely used feature selection technique, aims to ascertain the most relevant features that contribute drastically to the classification task. By incorporating this method, we pinpointed the most pertinent features related to diabetes detection, facilitating the generation of a more precise classifier.

The capsule network model, a relatively new deep learning architecture, was employed to leverage the selected features for diabetes classification. The accuracy, recall, F1 score, and class-wise precision and recall metrics were thoroughly investigated to determine the efficacy of the suggested strategy. The above average precision, recall, and F1 scores obtained indicate that the combination of Fisher score feature selection and capsule networks leads to accurate and reliable diabetes detection. One notable finding of this work is the high accuracy and recall values achieved by the proposed architecture. The high accuracy value depicts that the model efficaciously classified a significant proportion of both positive and negative cases of diabetes. This is vital in curtailing the risk of misdiagnosis, ensuring that patients receive suitable treatment and care. The class-wise precision and recall values further accentuate the proposed method' ability to suitably place occurrences of both classes, stressing its robustness in handling diverse cases of diabetes.

The kappa score, a statistical measure of agreement, gave supplementary evidence of the proposed method's performance. The high kappa score obtained (0.87) indicates substantial agreement between the algorithm's predictions and the actual diabetes labels, further strengthening the reliability and effectiveness of the proposed approach. The MSE, RMSE, and MAE values were calculated to assess the model's predictive accuracy. The small values obtained for these metrics (MSE: 0.058, RMSE: 0.24, MAE: 0.058) suggest that the proposed model's predictions are adjacent to the true values, demonstrating its capability to accurately estimate presence of diabetes. Table 2, comprises the summary of the previous proposed models and their performance evaluation metrics score.

http://dx.doi.org/10.36596/jcse.v4i1.708

Table 2. Comparison with Latest Contemporary Findings

Previous Works	Feature Selection Method	Model Used	Best Performing Model	Performance Evaluation Metrics Score
[14]	Kernel entropy component analysis	RF, LR, GNB, SVM, LDA, KNN, EGB, DT	RF	Accuracy = 96.75%, Precision = 99.64%, ROC value = 99%
[20]	-	DT, LR, DA, SVM, KNN, EN	LR	Accuracy = 77.9%
[23]	Greedy Stepwise Search	MLP, KM, LR, RF	MLP	Accuracy = 85.153%
[24]	-	SVM, KNN, GB, DT, RD, LR	RF	Accuracy = 84%, Precision = 83%, Recall = 76%, F1-score = 80%, ROC/AUC Score = 80%
[25]	-	GB, AB, XGB, NN, SVM, RF, SNN, SSVM, SRF	AB, RF, SVM	Accuracy = 98%, Recall = 98%, F1-Score = 98%
[28]	-	DT	DT	Accuracy = 78.1768%
[29]	Scatter Matrix	DT, ID3	ID3	Accuracy = 80%
[33]	Recursive Feature Elimination Genetic Algorithm, Burota Package.	DT	DT	Accuracy = 74.48%
[34]	-	RF, C4.5, RT, SC, BT, SVM	RF	Error Rate = 0.21
[35]	-	RF	RF	Sensitivity = 91.40%, Accuracy = 94.37
[36]	Minimal Redundancy Maximal Relevance	GB, SVM, AB, RF	RF	Accuracy = 99.35%
[37]	-	DT, SVM, NB	NB	Accuracy = 76.30%, Precision = 75.9%, Recall = 76.3%, F-Measure = 76%, ROC Score = 81.9%
[39]	Genetic Algorithm	NB	NB	Accuracy = 88.65%
[39]	Restricted Boltzmann machine	DT	DT	Accuracy = 80%, Kappa score = 0.5046, MAE = 0.35.71, RMSE = 0.434, RAE = 81.11%, RRSE = 92.65%,
[40]	-	NB, KNN, LR, RF	KNN	Accuracy = 78.57%, Precision = 87%, Specificity = 72%
[41]	Variational Autoencoder, Sparse Autoencoder	CNN	CNN	Accuracy = 92.37%
[42]	Pearson's Correlation Analysis	GNB, LSVM, RF	LSVM	Accuracy = 78.39%, Precision = 71.55%
[43]	Correlation-Based Feature Selection	ESVM, DNN	ESVM	Accuracy = 98.45%, RMSE = 0.568257571, R^2 = 0. 476923077, MAE = 0. 322916667, MSE = 0. 322916667
[44]	-	LR, NN, DJ, DT	DT	Accuracy = 80.2%, Precision = 73%, Recall = 62.9%, F1-Score = 67.6%
[45]	-	LR, NB, KNN, DT, RF, NN	NN	Accuracy = 99.15%, F1-Score = 99.63%, Precision = 99.27%, Recall = 100%, Specificity = 98.81%
[46]	Augmented Vector Foot	LSTM, 1DCNN, DECG, T, CNN16	1DCNN	Accuracy = 95.03%
[47]	-	NB, LR, RF	RF	Accuracy = 99% Precision = 99%, Recall = 99%, F-Measure = 98%

http://dx.doi.org/10.36596/jcse.v4i1.708

5. Conclusion

The research involves a comprehensive analysis of the Fisher score method to classify the most informative features for diabetes detection. These curated features are then employed to train a capsule network model, which is assessed using a dataset encompassing clinical features and subsequent diabetes labels. To determine the effectiveness of the suggested method, performance evaluation criteria such as accuracy, precision, recall, and F1 score are thoroughly studied. The results reveal that the fusion of Fisher score feature selection and capsule networks generates promising results in diabetes detection, as evidenced by the high accuracy and recall values. Furthermore, the algorithm correctly classifies both positive and negative cases of diabetes, dropping the possibility of misdiagnosis. By exploiting the Fisher score feature selection method and capsule networks, this work contributes to advancing diabetes detection and improving the accuracy and efficiency of diabetes detection.

The outcomes of this work demonstrate that the amalgamation of Fisher score feature selection and capsule networks holds great potential for augmenting diabetes detection. The proposed method effectively captures the relevant information necessary for accurate classification, leading to improved diagnostic accuracy and efficiency. The findings of this work augment the improvement of diabetes detection methods, proposing potential for individualized treatment options and better patient outcomes. While the outcomes of this work are encouraging, further reconnaissance and refinement of these approaches is warranted. Future research could reconnoiter the generalizability of the proposed methodology by validating it with larger and more diverse datasets. The integration of other relevant clinical features and the consideration of external factors, such as way of life and genetic evidence, could further uplift the algorithm's performance. Continued expansion and improvement of these procedures have the capability to meaningfully improve diabetes diagnosis, ultimately leading to more personalized and effective treatment strategies for individuals with diabetes.

Reference

[1] V. P, N. swamy P, and V. P, "Review on diabetes mellitus," *Journal of Innovations in Applied Pharmaceutical Science (JIAPS)*, pp. 24–27, Jan. 2022, doi: 10.37022/jiaps.v7i1.273.

[2] N. G. Forouhi and N. J. Wareham, "Epidemiology of diabetes," *Medicine*, vol. 38, no. 11, pp. 602–606, Nov. 2010, doi: 10.1016/j.mpmed.2010.08.007.

[3] M. Shivashankar, "A BRIEF OVERVIEW OF DIABETES."

[4] A. D. Wright, "What is diabetes?," in *Diabetic Retinopathy: Screening to Treatment (Oxford Diabetes Library)*, Oxford University Press, 2020, pp. 1–6. doi: 10.1093/med/9780198834458.003.0001.

[5] C. T. Wasnik, "Machine Learning Techniques: A Review," *Int J Res Appl Sci Eng Technol*, vol. 7, no. 5, pp. 3091–3095, May 2019, doi: 10.22214/ijraset.2019.5510.

[6] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science (1979)*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.

[7] T. Dietterich, "Machine learning," *ACM Comput Surv*, vol. 28, no. 4es, p. 3, Dec. 1996, doi: 10.1145/242224.242229.

[8] P. Singh, "Supervised Machine Learning," in *Learn PySpark*, Berkeley, CA: Apress, 2019, pp. 117–159. doi: 10.1007/978-1-4842-4961-1_6.

[9] K. Sindhu Meena and S. Suriya, "A Survey on Supervised and Unsupervised Learning Techniques," in *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*, Cham: Springer International Publishing, 2020, pp. 627–644. doi: 10.1007/978-3-030-24051-6_58.

[10] R. J. Tallarida and R. B. Murray, "Linear Regression I," in *Manual of Pharmacologic Calculations*, New York, NY: Springer New York, 1987, pp. 10–13. doi: 10.1007/978-1-4612-4974-0_4.

http://dx.doi.org/10.36596/jcse.v4i1.708

[11] S. Suthaharan, "Support Vector Machine," 2016, pp. 207–235. doi: 10.1007/978-1-4899-7641-3_9.
[12] L. Breiman, "Machine Learning," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[13] P. Cichosz, "Naïve Bayes classifier," in *Data Mining Algorithms*, Chichester, UK: John Wiley & Sons, Ltd, 2015, pp. 118–133. doi: 10.1002/9781118950951.ch4.
[14] P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 17, no. 1, p. 102690, Jan. 2023, doi: 10.1016/j.dsx.2022.102690.
[15] B. Everitt, "Cluster analysis," *Qual Quant*, vol. 14, no. 1, pp. 75–100, Jan. 1980, doi: 10.1007/BF00154794.
[16] G. Roșu and T. F. Șerbănuță, "An overview of the K semantic framework," *J Log Algebr Program*, vol. 79, no. 6, pp. 397–434, Aug. 2010, doi: 10.1016/j.jlap.2010.03.012.
[17] A. Wasilewska, "APRIORI Algorithm Lecture Notes."
[18] H. U. R. Siddiqui *et al.*, "Enhancing Cricket Performance Analysis with Human Pose Estimation and Machine Learning," *Sensors*, vol. 23, no. 15, p. 6839, Aug. 2023, doi: 10.3390/s23156839.
[19] S. Nasim, M. S. Almutairi, K. Munir, A. Raza, and F. Younas, "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics," *IEEE Access*, vol. 10, pp. 97610–97624, 2022, doi: 10.1109/ACCESS.2022.3205587.
[20] A. Al-Zebari and A. Sengur, "Performance Comparison of Machine Learning Techniques on Diabetes Disease Detection," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, IEEE, Nov. 2019, pp. 1–4. doi: 10.1109/UBMYK48245.2019.8965542.
[21] T. Sharma and M. Shah, "A comprehensive review of machine learning techniques on diabetes detection," *Vis Comput Ind Biomed Art*, vol. 4, no. 1, p. 30, Dec. 2021, doi: 10.1186/s42492-021-00097-7.
[22] S. Gujral, "Early Diabetes Detection using Machine Learning: A Review," *IJIRST-International Journal for Innovative Research in Science & Technology*|, vol. 3, no. 10, 2017, [Online]. Available: www.ijirst.org
[23] S. Y. Rubaiat, M. M. Rahman, and Md. K. Hasan, "Important Feature Selection & Accuracy Comparisons of Different Machine Learning Models for Early Diabetes Detection," in *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, IEEE, Dec. 2018, pp. 1–6. doi: 10.1109/CIET.2018.8660831.
[24] R. Katarya and S. Jain, "Comparison of Different Machine Learning Models for diabetes detection," in *2020 IEEE International Conference on Advances and Developments in Electrical and Electronics Engineering (ICADEE)*, IEEE, Dec. 2020, pp. 1–5. doi: 10.1109/ICADEE51157.2020.9368899.
[25] M. S. Tahsin, M. Jobayer, Md. B. U. Antor, M. Islam, F. F. Raisa, and Md. A. H. Shaikat, "Predictive Analysis & Brief Study of Early-Stage Diabetes Using Multiple Classifier Models," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, Jan. 2022, pp. 0203–0207. doi: 10.1109/CCWC54503.2022.9720736.
[26] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, Elsevier, 2018, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.
[27] S. Aronoff, "Classification Accuracy: A User Approach."
[28] A. A. Al Jarullah, "Decision tree discovery for the diagnosis of type II diabetes," in *2011 International Conference on Innovations in Information Technology*, IEEE, Apr. 2011, pp. 303–307. doi: 10.1109/INNOVATIONS.2011.5893838.
[29] J. Han, J. C. Rodriguez, and M. Beheshti, "Discovering Decision Tree Based Diabetes Prediction Model," 2009, pp. 99–109. doi: 10.1007/978-3-642-10242-4_9.
[30] F. Li and Y. Yang, "Analysis of recursive feature elimination methods," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA: ACM, Aug. 2005, pp. 633–634. doi: 10.1145/1076034.1076164.
[31] R. Leardi, R. Boggia, and M. Terrile, "Genetic algorithms as a strategy for feature selection," *J Chemom*, vol. 6, no. 5, pp. 267–281, Sep. 1992, doi: 10.1002/cem.1180060506.

http://dx.doi.org/10.36596/jcse.v4i1.708

[32] M. B. Kursa and W. R. Rudnicki, "Feature Selection with the Boruta Package," *J Stat Softw*, vol. 36, no. 11, 2010, doi: 10.18637/jss.v036.i11.

[33] J. Sadhasivam, V. Muthukumaran, J. Thimmia Raja, R. B. Joseph, M. Munirathanam, and J. M. Balajee, "Diabetes disease prediction using decision tree for feature selection," *J Phys Conf Ser*, vol. 1964, no. 6, p. 062116, Jul. 2021, doi: 10.1088/1742-6596/1964/6/062116.

[34] S. Benbelkacem and B. Atmani, "Random Forests for Diabetes Diagnosis," in *2019 International Conference on Computer and Information Sciences (ICCIS)*, IEEE, Apr. 2019, pp. 1–4. doi: 10.1109/ICCISci.2019.8716405.

[35] N. ZAABOUB and A. DOUIK, "Early Diagnosis of Diabetic Retinopathy using Random Forest Algorithm," in *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, IEEE, Sep. 2020, pp. 1–5. doi: 10.1109/ATSIP49331.2020.9231795.

[36] P. Ghosh, S. Azam, A. Karim, M. Hassan, K. Roy, and M. Jonkman, "A Comparative Study of Different Machine Learning Tools in Detecting Diabetes," *Procedia Comput Sci*, vol. 192, pp. 467–477, 2021, doi: 10.1016/j.procs.2021.08.048.

[37] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput Sci*, vol. 132, pp. 1578–1585, 2018, doi: 10.1016/j.procs.2018.05.122.

[38] O. Somantri, "An Optimize Weights Naïve Bayes Model for Early Detection of Diabetes," *Telematika*, vol. 15, no. 1, Feb. 2022, doi: 10.35671/telematika.v15i1.1307.

[39] T. P. Kamble, "Diabetes Detection using Deep Learning Approach", [Online]. Available: www.ijirst.org

[40] L. V. R. Kumari, P. Shreya, M. Begum, T. P. Krishna, and M. Prathibha, "Machine Learning based Diabetes Detection," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Jul. 2021, pp. 1–5. doi: 10.1109/ICCES51350.2021.9489058.

[41] M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation," *Comput Methods Programs Biomed*, vol. 202, p. 105968, Apr. 2021, doi: 10.1016/j.cmpb.2021.105968.

[42] T. A. Assegie and S. Nair, "The Performance Of Different Machine Learning Models On Diabetes Prediction," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 9, p. 1, 2020, [Online]. Available: www.ijstr.org

[43] P. Nagaraj and P. Deepalakshmi, "Diabetes Prediction Using Enhanced SVM and Deep Neural Network Learning Techniques," *International Journal of Healthcare Information Systems and Informatics*, vol. 16, no. 4, pp. 1–20, Sep. 2021, doi: 10.4018/IJHISI.20211001.oa25.

[44] C.-Y. Chou, D.-Y. Hsu, and C.-H. Chou, "Predicting the Onset of Diabetes with Machine Learning Methods," *J Pers Med*, vol. 13, no. 3, p. 406, Feb. 2023, doi: 10.3390/jpm13030406.

[45] R. M. S, S. H. S, and S. M. S, "Machine Learning Algorithms for the Detection of Diabetes," *International Research Journal of Engineering and Technology*, 2021, [Online]. Available: www.irjet.net

[46] A. R. Kulkarni *et al.*, "Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram," *BMJ Innov*, vol. 9, no. 1, pp. 32–42, Jan. 2023, doi: 10.1136/bmjinnov-2021-000759.

[47] M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, "Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques," 2020, pp. 113–125. doi: 10.1007/978-981-13-8798-2_12.

[48] N. Xiao, Z. Wang, Y. Huang, F. Daneshgari, and G. Liu, "Roles of Polyuria and Hyperglycemia in Bladder Dysfunction in Diabetes," *Journal of Urology*, vol. 189, no. 3, pp. 1130–1136, Mar. 2013, doi: 10.1016/j.juro.2012.08.222.

[49] L. Ahmadi and M. B. Goldman, "Primary polydipsia: Update," *Best Pract Res Clin Endocrinol Metab*, vol. 34, no. 5, p. 101469, Sep. 2020, doi: 10.1016/j.beem.2020.101469.

[50] E. A. Homan, M. V. Reyes, K. T. Hickey, and J. P. Morrow, "Clinical Overview of Obesity and Diabetes Mellitus as Risk Factors for Atrial Fibrillation and Sudden Cardiac Death," *Front Physiol*, vol. 9, Jan. 2019, doi: 10.3389/fphys.2018.01847.

[51] H. Andersen, "Motor dysfunction in diabetes," *Diabetes Metab Res Rev*, vol. 28, pp. 89–92, Feb. 2012, doi: 10.1002/dmrr.2257.

http://dx.doi.org/10.36596/jcse.v4i1.708

[52] L. Hernandez and E. Briese, "Analysis of diabetic hyperphagia and polydipsia," *Physiol Behav*, vol. 9, no. 5, pp. 741–746, Nov. 1972, doi: 10.1016/0031-9384(72)90044-3.

[53] S. Kumar, A. J. Costello, and P. G. Colman, "Fournier's gangrene in a man on empagliflozin for treatment of Type 2 diabetes," *Diabetic Medicine*, vol. 34, no. 11, pp. 1646–1648, Nov. 2017, doi: 10.1111/dme.13508.

[54] J. Xue, F. Min, and F. Ma, "Research on Diabetes Prediction Method Based on Machine Learning," *J Phys Conf Ser*, vol. 1684, p. 012062, Nov. 2020, doi: 10.1088/1742-6596/1684/1/012062.

[55] J. A. Vazquez and J. D. Sobel, "FUNGAL INFECTIONS IN DIABETES," *Infect Dis Clin North Am*, vol. 9, no. 1, pp. 97–116, Mar. 1995, doi: 10.1016/S0891-5520(20)30642-5.

[56] D. H. C. Surridge *et al.*, "Psychiatric Aspects of Diabetes Mellitus," *British Journal of Psychiatry*, vol. 145, no. 3, pp. 269–276, Sep. 1984, doi: 10.1192/bjp.145.3.269.

[57] R. Blakytny and E. Jude, "The molecular biology of chronic wounds and delayed healing in diabetes," *Diabetic Medicine*, vol. 23, no. 6, pp. 594–608, Jun. 2006, doi: 10.1111/j.1464-5491.2006.01773.x.

[58] J. Ma, "Machine Learning in Predicting Diabetes in the Early Stage," in *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, IEEE, Oct. 2020, pp. 167–172. doi: 10.1109/MLBDBI51377.2020.00037.

[59] G. B. Salsich, M. Brown, and M. J. Mueller, "Relationships Between Plantar Flexor Muscle Stiffness, Strength, and Range of Motion in Subjects With Diabetes-Peripheral Neuropathy Compared to Age-Matched Controls," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 30, no. 8, pp. 473–483, Aug. 2000, doi: 10.2519/jospt.2000.30.8.473.

[60] L.-H. Su, L.-S. Chen, S.-C. Lin, and H.-H. Chen, "Association of Androgenetic Alopecia With Mortality From Diabetes Mellitus and Heart Disease," *JAMA Dermatol*, vol. 149, no. 5, p. 601, May 2013, doi: 10.1001/jamadermatol.2013.130.

[61] S. Verma and M. E. Hussain, "Obesity and diabetes: An update," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 11, no. 1, pp. 73–79, Jan. 2017, doi: 10.1016/j.dsx.2016.06.017.

[62] M. Gan and L. Zhang, "Iteratively local fisher score for feature selection," *Applied Intelligence*, vol. 51, no. 8, pp. 6167–6181, Aug. 2021, doi: 10.1007/s10489-020-02141-0.

[63] Q. Gu, Z. Li, and J. Han, "Generalized Fisher Score for Feature Selection."

[64] J. Li *et al.*, "A Survey on Capsule Networks: Evolution, Application, and Future Development," in *2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, IEEE, Dec. 2021, pp. 177–185. doi: 10.1109/HPBDIS53214.2021.9658349.

[65] R. Shi and L. Niu, "A brief survey on Capsule Network," in *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE, Dec. 2020, pp. 682–686. doi: 10.1109/WIIAT50758.2020.00103.

[66] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules."

[67] J. Tao, X. Zhang, X. Luo, Y. Wang, C. Song, and Y. Sun, "Adaptive Capsule Network," *Computer Vision and Image Understanding*, vol. 218, p. 103405, Apr. 2022, doi: 10.1016/j.cviu.2022.103405.

[68] M. Kwabena Patrick, A. Felix Adekoya, A. Abra Mighty, and B. Y. Edward, "Capsule Networks – A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 1, pp. 1295–1310, Jan. 2022, doi: 10.1016/j.jksuci.2019.09.014.

[69] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.

[70] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning - ICML '06*, New York, New York, USA: ACM Press, 2006, pp. 233–240. doi: 10.1145/1143844.1143874.

[71] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," 2006, pp. 1015–1021. doi: 10.1007/11941439_114.

[72] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit Lett*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

http://dx.doi.org/10.36596/jcse.v4i1.708

[73] L. Gonçalves, A. Subtil, M. R. Oliveira, and P. de Z. Bermudez, "ROC curve estimation: An overview," *REVSTAT-Statistical Journal*, vol. 12, no. 1, 2014.

[74] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf Process Manag*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

[75] T. R. Shultz *et al.*, "Confusion Matrix," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 209–209. doi: 10.1007/978-0-387-30164-8_157.

[76] B. E. Wampold, "Kappa as a measure of pattern in sequential data," *Qual Quant*, vol. 23, no. 2, pp. 171–187, Jun. 1989, doi: 10.1007/BF00151902.

[77] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ Psychol Meas*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: 10.1177/001316446002000104.

[78] Zhou Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Process Mag*, vol. 26, no. 1, pp. 98–117, Jan. 2009, doi: 10.1109/MSP.2008.930649.

[79] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geosci Model Dev*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/gmd-7-1247-2014.

[80] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim Res*, vol. 30, pp. 79–82, 2005, doi: 10.3354/cr030079.

http://dx.doi.org/10.36596/jcse.v4i1.708