Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Optimizing Multiclass Banking Customer Segmentation: A Dual Approach using Transformer Model and Advanced Optimizers

Sufaira Shamsudeen[1*], K. Ranjith Singh [2]

[1]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India,

Email: sufaira@mesmarampally.org.

*Department of Computer Applications, MES College Marampally, Aluva, Ernakulam, Kerala, India, Email: sufaira.mes@gmail.com.

[2]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India,

Email: ranjithsingh.koppaiyan@kahedu.edu.in.

Abstract

Abstract: Financial institutions and banking sectors, enhance customer satisfaction, improve customer retention rates, and optimize revenue of the business through customer segmentation. Customer segmentation is a key factor to consider in the business; today's business world is struggling to handle their customers in an effective way. As the business can create a competitive edge by focusing more on customer than the product. While identifying the loyal customers, they can offer more product offerings, tailored service, and improve the business revenue. Consequently, for the effective customer segmentation, the artificial intelligence, machine learning and deep learning techniques played a major role in the present day. Through deep learning techniques the model learned the customer behavior in many aspects and categorized, however, this architecture proved reliable models for customer segmentation. To handle the complexity and diversity of the customer information, traditional machine learning approaches often struggle. Hence, the research explores the deep learning models – Artificial Neural Networks (ANN), Gated Recurrent Units (GRU), and Transformer Networks (TN), for multiclass customer classification. ANN is a powerful technique to learn effectively the data with complex non-linear pattern, while GRU are effective in handling the sequential data, and TN are excellent for self-attention mechanisms. Based on the metrics such as accuracy, precision, recall, and f1-score, all models performs well. Among these, transformer model excelled others in capturing complex data patterns, and providing superior segmentation. The study explores the efficacy of advanced deep learning architectures for better customer prediction and segmentation; in that way predict the profitable customer and an upturn in the economy.

Keywords: Deep Learning; Artificial Neural Network; Gated Recurrent Unit; Transformer Network; Machine Learning

1. Introduction

In the economical banking business industry, utilizing the customer data for better growth and decision-making is very essential. Customer-centric approaches have become crucial for banks to maintain a competitive edge in the business world. In this context, to drive for success, particularly in financial institutions and the banking sector, customer segmentation played a pivotal role. Customer segmentation [1, 2], as a means to enhance customer satisfaction, improve retention rates, and optimize revenue generation [1], partitions the customers into distinct groups based on their preferences, demographics, and financial behaviors. Therefore, data acquisition from the relevant sources, preprocessing the raw data, and applying some techniques to these data can lead to valuable insights, thereby increasing the decision-making. To deal with these challenges, artificial intelligence (AI) [1] techniques like machine learning (ML) [1] and deep learning (DL) [1] architecture have emerged as an improved solution.

A vast amount of heterogeneous data in the bank, such as transaction history, loan details, customer information, and account details, leads to more complexity to explore; as a result systematically assessing the data, enables the banks to target profitable customers and tailor the products and services efficiently [1], so enhancing the revenue. In the 21st century, the big data revolution [3, 4, 5, 6] has significantly affected the banking firm due to the vast amount of data, which is valuable for economic growth and has been mounting up for decades. Big data analytics helps banks to analyse the customers' behaviour and preferences, create precise customer segments and profile, perform sentimental analysis, and enhance security protocols and fraudulent detection in banking [5]. Effective use of this information [6] explores some patterns that show understandings of the financial transactions; in this manner, it provides more transaction movements, enhanced customer preferences [3], and identifies fraud [6] and unauthorized access [2, 6]. Taking on the sophisticated architectures [3], easily interpret the patterns of data from the various sources [7] solve the problem. The diverse nature of the data is recycled to make a data-driven decision-making system and, therefore, make their business more successful [7].

AI and ML techniques adopted in many financial industries, especially in the banking sector for the risk management and other various operations [8]. Because of the diverse nature of the data and large size, deep learning architecture helps to predict and classify the customer data more precisely. It is one of the emerging technologies; recently, the banking sector adopted this technique to manipulate the various operations efficiently [9]. Based on the review, there is no detailed study done using the advanced deep learning techniques. As we dive into the fast-evolving technology in the 21st century of AI, the complexity of the data grown more and the deep learning approach employed to handle the complex patterns. This transformative field enable the system to independently discover the complex patterns and build a decision support system from the unstructured huge amount of data. DL applied in many areas such as medical diagnosis system, natural language processing, and business data analytics etc., pave a future system, and observe, learn, and transform the data autonomously.

In this article, deep learning architectures such as Artificial Neural Network (ANN) implementation for customer segmentation and advanced learning architectures like Gated Recurrent Unit (GRU) and Transformer Network (TN) gaged for better customer segmentation. These models excel in capturing the data patterns very effectively and have shown a remarkable ability to classify the customers into five distinct classes, thus delivering reliable customer classification. By adopting the evaluation metrics such as accuracy, precision, recall, and f1-measure, the conclusion of the research by showing the superiority of the transformer network for providing precise segmentation. The exploration focus on the article is as follows:

1 A comprehensive study of various classification models counted in.
2 Explain in detail the deep learning architecture ANN, GRU, and Transformer Network, which have implemented for better customer segmentation.
3 Explore the performance of each model by examining various evaluation metrics.
4 Gain knowledge to the researchers about deep learning architectures used for customer segmentation.
5 This article provides future research walks in this domain.

The remainder of the article systematized as follows: the literature reviews presented in section II. The section III, explains the proposed framework adopted for multiclass classification. Proposed methods described in Section IV. Section V outlines the results, followed by the evaluation metrics used for the comparison. Challenges and future scope discussed in Section VI. Last of all, the conclusion of the article discussed in section VII.

2. Related Surveys

Several articles have been referenced based on the models as stated above. Md et al. [1] performed study on the customer segmentation in the banking industry and its significance for the economic growth of business using the machine learning architecture. Gankidi et al. [2] various classification models surveyed such as supervised and unsupervised machine learning algorithms for multiclass customer segmentation. A huge value of heterogeneous data from various sources is transforming the business into potential revenue; hence, the evolution of big data analytics and utilizing it in an effective manner can assess and manage the risk and figure out the patterns explained in detail by Dicuonzo et al. [3]. More et al. [4] focused on the advantages of big data analytics in the banking sector, such as fraud detection, customer segmentation, managing risk, and enhancing revenue based on historical data to predict the future. Also focus on some sophisticated algorithms implemented for the various operations in the bank Utkarsh Srivastava et al. [5], Gupta et al. [6]. Milojević et al. [8] explains the significance of artificial intelligence, machine learning, and deep learning strategies in the financial sectors, especially in banking. The author emphasizes the broad range of AI in the banking sector, and ML, a subset of AI, has been employed to manage the risk in banking. Over the past two decades, DL, a refined architecture of ML, realized its strength in the banking industry by addressing various challenges, Milosevic et al. [8]. Moreover, big data analytics, an analysis of vast amounts of data, became popular in the economic world under the influence of sophisticated intelligence techniques such as AI, ML, and especially DL. Hassani et al. [9] cited the emerging technology Deep Learning (DL), data-driven technology has significant advancement in the banking sector in recent years. Shoaib et al. [10], discussed various deep learning approaches for multiclass classification.

3. Materials And Methods

3.1. Dataset

The accuracy of classification depends on the dimensions and features of the dataset. However, finding the real time data from the bank was very tedious task. The main challenge in collecting the data from the bank is they are not willing to explore the customer details due to confidentiality. In this work, the dataset with more than one lakh of records, with 43 features and one target attribute, collected from Indian Bank are used. Due to privacy concerns, the name of the bank will not be disclosed. The real dataset is probably messy and need data pre-processing such as analysing the data, denoising, imputation, data reduction, integer encoding categorical variables, oversampling, normalization, splitting dataset into training and test sets and feature scaling. The algorithm is then trained using the training set, then evaluated using cross-validation techniques and other non-parametric measures.

The bank customers are classified into five categories – Class0: Outstanding Customers, Class1: Excellent Customers, Class2: Good Customers, Class3: Satisfactory Customers, and Class4: Bad Customers. The Class0 and Class1 are taken as the profitable Customer, i.e., the Outstanding Customers and Excellent Customers, based on their credit history and thereby these classes of customers are cost-effective customer for the bank.

3.2. Classification Models

Classification is a supervised machine learning approach [11, 12, 13] is to build labelled classes based on the predictor features. Many applications like image processing, customer churn prediction [14, 15], credit card fraud detection [16, 17], medical diagnosis [18, 19, 20], spam detection [21] etc. are done using classification approach. Customer classification is a significant application used in many businesses and marketing world. In recent years, banking industry is focusing more on customers than the product, to hike the revenue of business. For better understanding of the customers' behavior and nature, an unconventional analytical tool [22] applied to segments the customers into distinct classes. In the proposed study, based on the demographic behavior and credit information the multiclass customer classification method have adopted. Mainly, classification are two – binary and multiclass classification [23]. In binary classification, classes are of two and in multiclass classification; many classes labelled based on the feature of customers.

In the proposed work, highlights classifying the bank customer data depends on the demographic and credit information into five distinct labelled classes using the deep learning architectures – ANN, GRU and TN. Based on the dataset size and features, TN model outperforms. ANN and GRU models results better classification but TN gives superior accuracy and other metrics value. Most of the surveys reviewed, noticed that the classification is implemented using ANN model in various fields. GRU and TN model is a novel approach, which demonstrates precise and efficient computation in complex data.

3.2.1. Artificial Neural Network

Machine learning model that can analyze the complex pattern and learn the complex relationship between input and output variables is Neural Network Model. Generally, they are regularly for classification and regression purpose. The accuracy measure of the model considered some features such as parameters taken, quality of training data, and the complexity of the data. In general, compared to the traditional machine learning models, neural network handle the complex problems. An architecture of neural network is defined as the number of artificial neurons (units) arranged in sequence of layers [24, 25]. One simple and common type of neural network, namely Multilayer Perceptron (MLP) [26, 27], which are commonly used for multiclass customer segmentation with simple dataset. Artificial Neural Network (ANN) [28, 29], have been executed to segment retail customer database for marketing purpose, by providing more homogenous segments. According to survey, ANN dense model can handle complex and large database very efficiently.

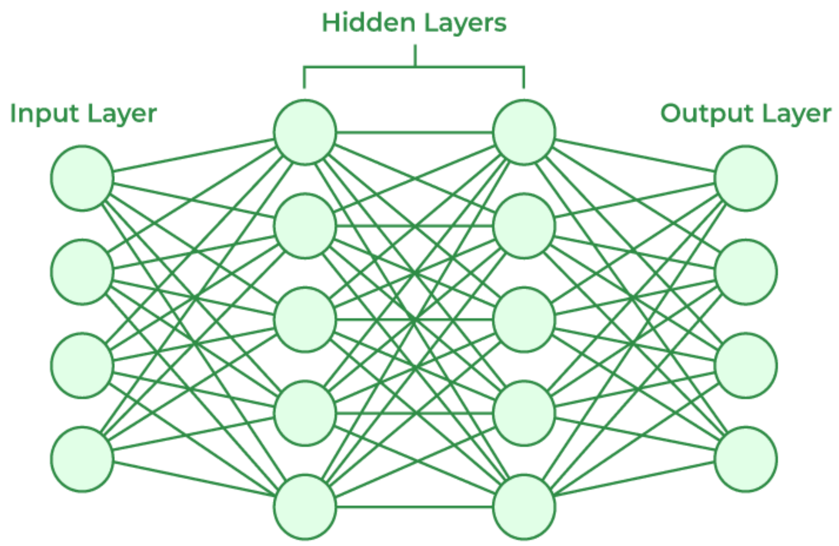


Fig 1: Neural Networks Architecture [29]

Keras dense layer neural network architecture is used here for classification purposes. Neural networks fundamentally rely on the dense layer [29], a deeply connected layer. The implementation of layer is principally feedforward backpropagation architecture. It consists of one input layer, multiple dense layers, i.e., hidden layers, and one output layer. Dense layer perform a linear transformation on its input, then apply the non-linear activation function in the dense layer, such as ReLU, Sigmoid, and Tanh, which is utilized to learn intricate patterns in data. The output of the model is generated after executing the activation function, which may be input for the subsequent layers and so on. Finally, the last layer produces the target predictions, i.e., the final output. Feedforward backpropagation neural architecture is a combination of forward propagation and backward propagation procedures to learn a model. Principally, it includes two steps: forward propagation and backward propagation. The architecture fed the input and estimated the output in forward propagation. In backward propagation, the weights and biases adjusted in accordance with the error between the predicted output and actual output is calculated. Hence, reduce the error for all the neurons and produce better network prediction.

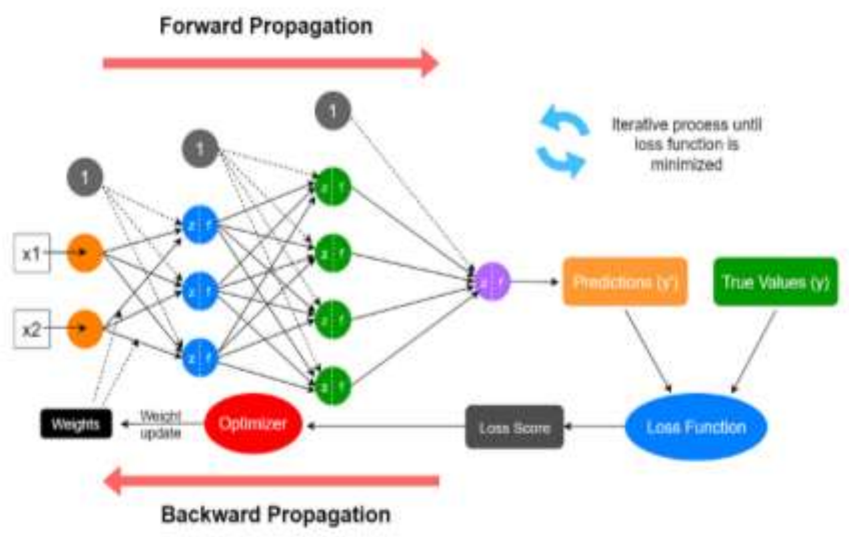


Fig 2: Learning process of Neural Networks [30]

3.2.2. Gated Recurrent Unit

Gated Recurrent Unit (GRU), a form of Recurrent Neural Network (RNN) architecture developed in 2014 [31, 32, 33], to handle sequential data. The structure of the sequential data may be vary such as time series, natural language, user behaviour, or encoded categorical data. GRU performs one data at a time, using the existing input and past output to updates the hidden state. The memory cell of GRU combines the input with the past hidden state. In the subsequent steps, active to update the hidden state. Compared to Long Short-Term Memory (LSTM), it is a variant and simplified version of LSTM. GRU merges long and short-term memory with the feature of single hidden state. Long Short-Term Memory (LSTM) with three gates, while GRU has memory cell comprises two gates: Update gate and Reset gate. How much of the past memory is discarded is controlled by reset gate, however the update gate defines how much of the past memory to retain.

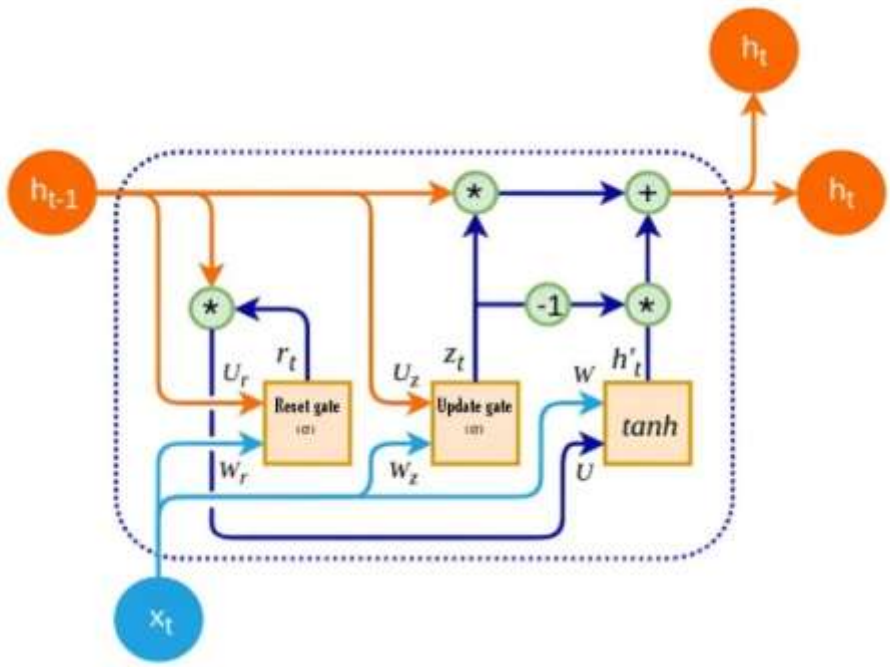


Fig 3. Structure of Gated Recurrent Unit Cell [32]

3.2.3. Transformer Network

Transformers [34, 35, 36] are not dependent on recurrence or convolutions, in contrast to earlier models like recurrent neural networks (RNNs) or convolutional neural networks (CNNs). Rather, they employ self-attention processes to record the connections among various items in a sequence, irrespective of their separation from one another. Transformers are very effective in managing long-range dependencies in data because of this. Transformers are more efficient than conventional RNNs, which process data sequentially, because of the parallel nature of attention computation. The transformer, nn advanced method in deep learning has shown to be incredibly successful in a variety of applications, ranging from natural language processing duties like translation and summarization to other domains like image processing and reinforcement learning.

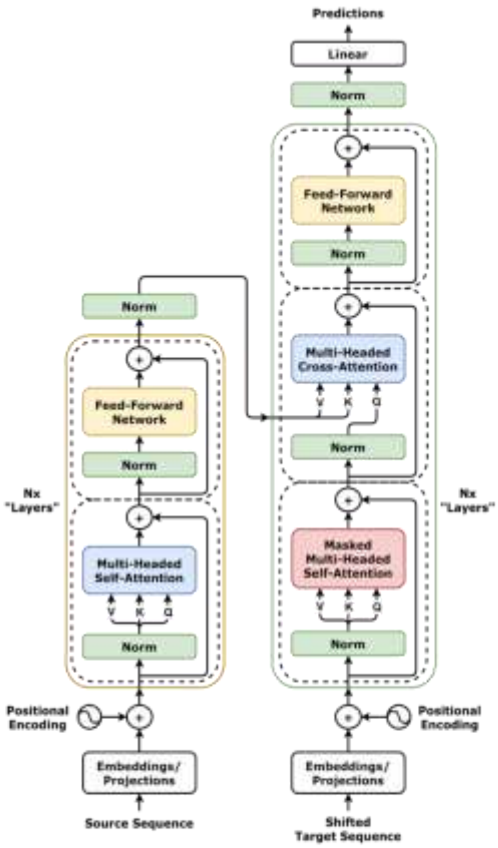


Fig 4. Standard Transformer Architecture [36]

3.3 Proposed Methodology

Fig. 5 illustrates the general framework of a proposed methodology. The study includes various steps for customer classification, i.e., data acquisition, data cleaning and preprocessing (data transformation), training and implementing multiclass classifiers, evaluation of performance indicators, and output visualization based on the metrics. Based on the visualization and metrics, model accurately chosen. The development of advanced technologies contributes to the economic growth of businesses across various sectors, including healthcare, banking, marketing, and security maintenance. Recently, business intelligence (BI), a decision support system, played a great role in the economic world [37]. Business intelligence places emphasis on the exploration of large volumes of data about the firm and its operations, then extracting the knowledge from the input data after processing [37].

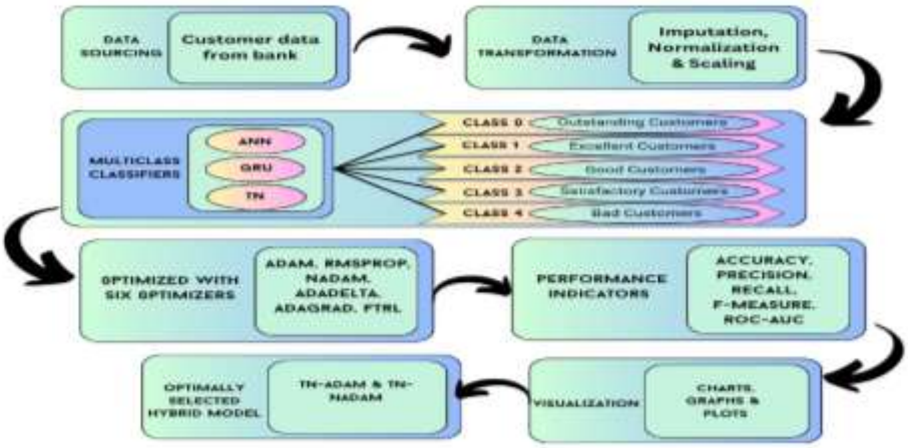


Fig 5. The framework of the Proposed Methodology

3.3.1. Data Sourcing

Finding the real time data from the bank was very tedious task. The main challenge in collecting the data from the bank is they are not willing to explore the customer details due to confidentiality. In this work, the dataset with more than one lakh of records, with 43 labelled input features, collected from Indian Bank are used. Due to privacy concerns, the name of the bank will remain confidential.

3.3.2. Data Transformation

After the data sourcing process, normalize the data using preprocessing procedures such as removing corrupt and missing values. The big amount of information from various sources, which may be unstructured or structured, in the database. For knowledge extraction from the database, the data transformation step is necessary. Unusual data may cause inconsistency in the output. To extend the quality of incoming data, significant steps in data preprocessing include cleaning of data, integration, reduction, transformation, discretization, and normalization. Ensuring the quality of the input data is essential cause for the better performance model. Tangible data may be noisy and contain missing values. To handle the missing values, a data imputation process needed. Imputation avoids dataset falsification and reduces model bias, thereby improving accuracy. The data cleaning method used to avoid the missing values and inconsistencies in the input data because most of the prediction algorithms do not predict the accurate result with the missing values. Usually missing values are imputed using mean, median, or mode methods. Merging data from different sources into a repository is a data integration process. Reducing the size and features of the data set by eliminating the replicated and irrelevant features is a data reduction procedure. After reducing the data attributes, normalize or standardize the data, that is, transform the data into an applicable form. This method is data transformation. Data transformation techniques are key for improving the performance of the classification models. If the data transformation procedure is completed, an appropriate classification model used here to classify, train, and evaluate data.

3.3.3. Proposed Multiclass Classifiers

The proposed study, focused on deep learning architectures—ANN, GRU, and TN models—trained for customer classification. The review of the papers shows that the ANN model is trained for classification in many fields. However, for the information, the GRU and TN models are novel approaches. Using these models, the dataset, which has 43 labelled input features, preprocessed and transformed to train and classify the customers into five distinct classes (Class 0 to Class 4).

3.3.3.1. Artificial Neural Network

Keras dense layer neural network architecture used here for classification purpose. Dense layer, a deeply connected layer, which is a fundamental component of neural networks. The implementation of dense layer is principally feedforward backpropagation architecture. ANN consist of one input layer and multiple hidden layer, and one output layer. In the proposed work, two dense layers are used, each with input layer consist of 43 labelled input features of the bank customer data, represented as input neuron, is associated to every output neuron, allowing for the capture of complex representation within the data are implemented and classify the data into five distinct classes. The structure of dense layer, it contains neuron, which receives input from the previous neurons. Each dense layer has its own parameters, weights and biases to learn the network during training. The activation function, ReLU used in the two hidden layers, Sigmoid and Softmax in the output layer for multiclass classification [29].

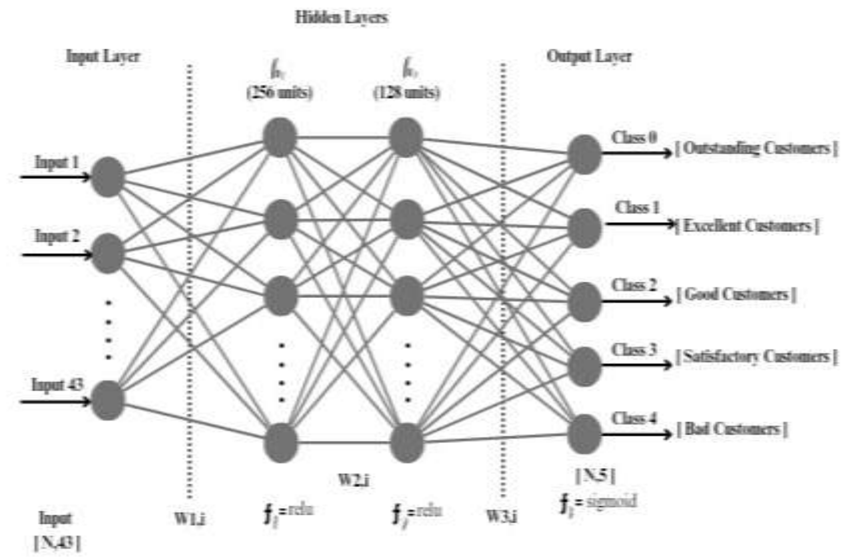


Fig 6. The layout of the Proposed ANN Model

3.3.3.2. Gated Recurrent Unit

Gated Recurrent Unit (GRU), is a simple architecture, and faster to train. The special feature of GRU is its hidden layer consist of two gates: update and reset gate, for the flow of information within the network [32]. In the proposed work, initialize the weights of the class, Weights (W_z, U_z, b_z, W_r, U_r, b_r, W_h, U_h, b_h), and then compute the update and reset gate.

[32, 33] Update Gate, $z_t = \sigma (W_z x_t + U_z h_{t-1} + b_z)$…………………………………… (1)

[32, 33] Reset Gate, $r_t = \sigma (W_r x_t + U_r h_{t-1} + b_r)$ ………..……………………………… (2)

To find the Hidden state, h_t in GRU, it follows a two-step process. The first step is to calculate the candidate hidden state. Then, measure final Hidden state.

[32, 33] Candidate Hidden State, $h'_t = \tanh (W_h x_t + U_h (r_t \odot h_{t-1}) + b_h)$…………….. (3)

Final,

[32, 33] Hidden State, $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h'_t$ ……………………………... (4)

Here, the value of the reset gate (r_t), is used to determine the previous hidden state (h_{t-1}) that is the parameter of candidate hidden state (h'_t). If $r_t = 1$, the entire information from the previous hidden state h_{t-1} is being measured, otherwise, i.e., $r_t = 0$, the information from the previous hidden state is completely ignored. Here, the process of computing memory cell gate, using hadamard product of the reset gate and previously hidden state.

In the GRU architecture, using two dense layers with ReLU activation function are computed, to enhance the model's ability to learn. The first dense layer contain 512 units, which allow the model to learn and the second layer further processes with reduced number of units to 256, to refine the learned features and produce better classification.

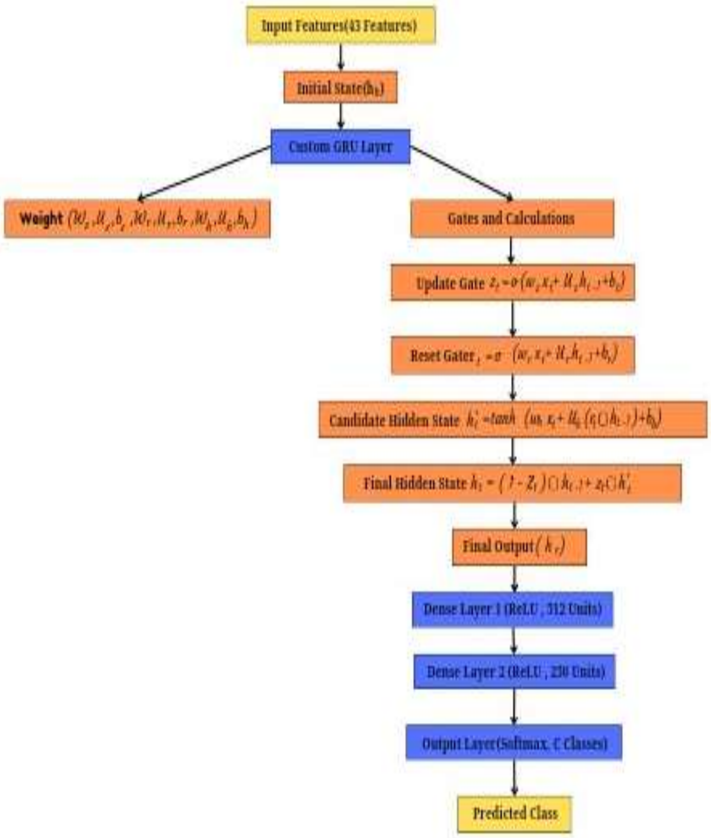


Fig 7. The layout of the Proposed GRU Model

3.3.3.3. Transformer

While considering the significance of input data sequence, the attention mechanism in transformer enables the model to prioritise certain segments of the input sequence when forming predictions. By using the multi-head self-attention mechanism in the transformer model, compute several attention distributions simultaneously.

An embedding layer converts each input token into a dense vector representation before adding positional encodings to preserve order information. A feed-forward neural network receives the output from the attention layers and applies non-linear transformations to improve the model's learning ability. During this process, layer normalization stabilizes training, enabling deeper architectures to converge more successfully, and residual connections are used to enhance gradient flow.

The Transformer can learn hierarchical representations of the input data because of the architecture's multiple stacked layers. The feed-forward network and attention mechanism make up each layer, which improves the model's ability to recognize complex patterns. The final step of processing yields the desired output, such as class probabilities in classification tasks, when a final linear transformation is combined with a softmax activation function.

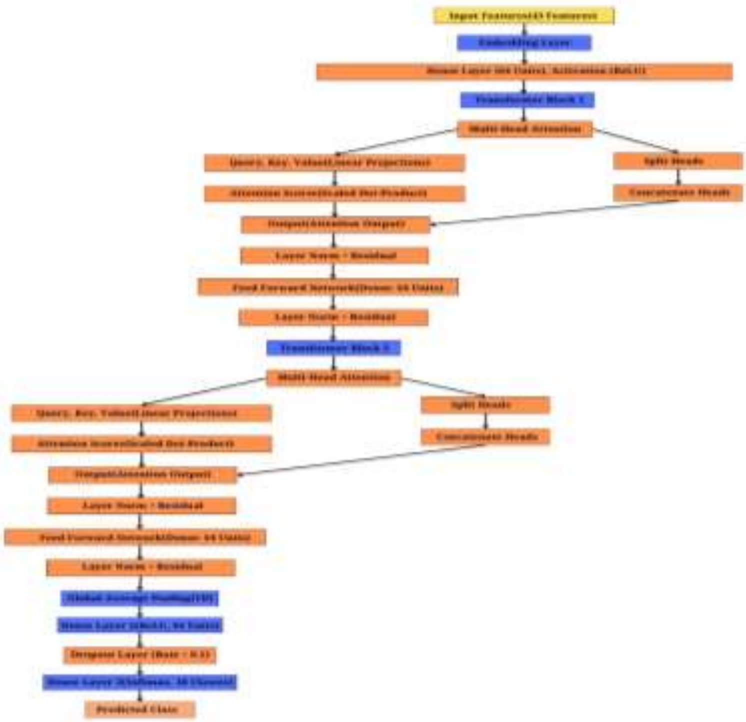


Fig 8. The layout of the Proposed TN Model

3.3.4. Performance Evaluation and Visualization

Upon the completion of classification, the performance of the model is evaluated using indicators such as accuracy, precision, recall, f1-score, and roc-auc. Finally, the results compared to figure out the best models by demonstrating the performance using visualization tools. Result visualization provides the perspective on the data and discover trends in the data. The results plotted using the different visualization tools. Charts, graphs and plots employed here to analyze and depict the output specifically. Visualization representations makes it easier to identify the patterns and make a comparison efficiently.

Table 1. Accuracy of ANN model with six Optimizers

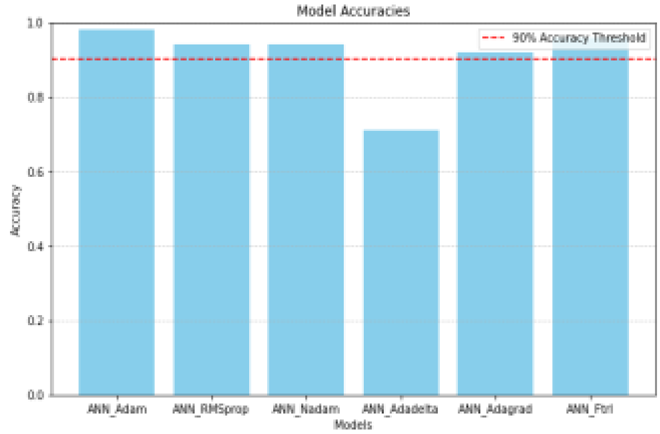
Models	**Accuracy**
ANN_Adam	0.98
ANN_RMSprop	0.94
ANN_Nadam	0.94
ANN_Adadelta	0.71
ANN_Adagrad	0.92
ANN_Ftrl	0.96

Table 2. Accuracy of GRU model with six Optimizers

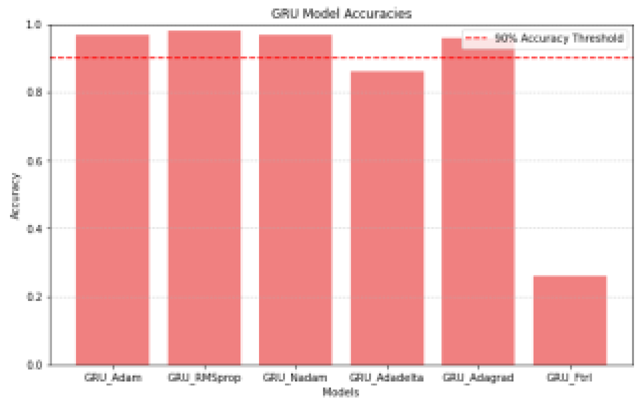
Models	**Accuracy**
GRU_Adam	0.97
GRU_RMSprop	0.98
GRU_Nadam	0.97
GRU_Adadelta	0.86
GRU_Adagrad	0.96
GRU_Ftrl	0.26

Table 3. Accuracy of TN model with six Optimizers

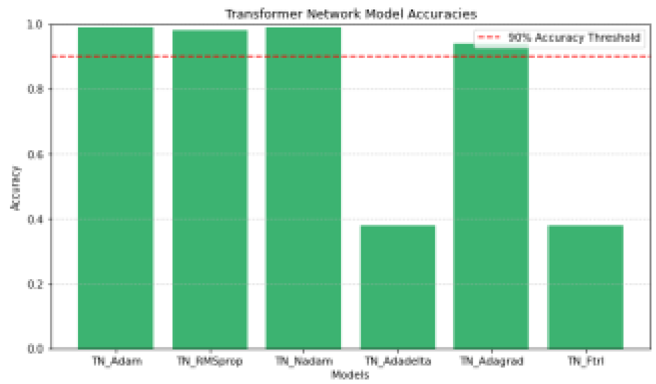
Models	Accuracy
TN_Adam	0.99
TN_RMSprop	0.98
TN_Nadam	0.99
TN_Adadelta	0.38
TN_Adagrad	0.94
TN_Ftrl	0.38



(a)

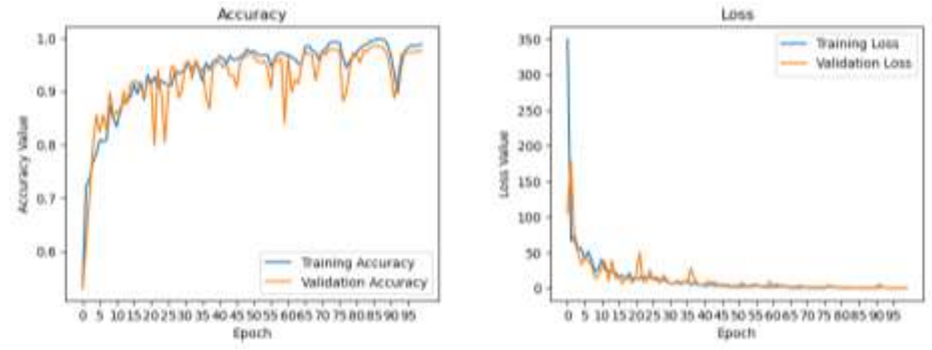


(b)

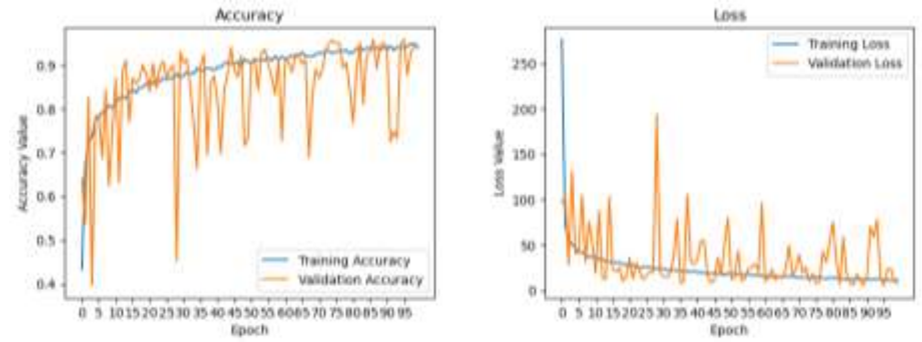


(c)

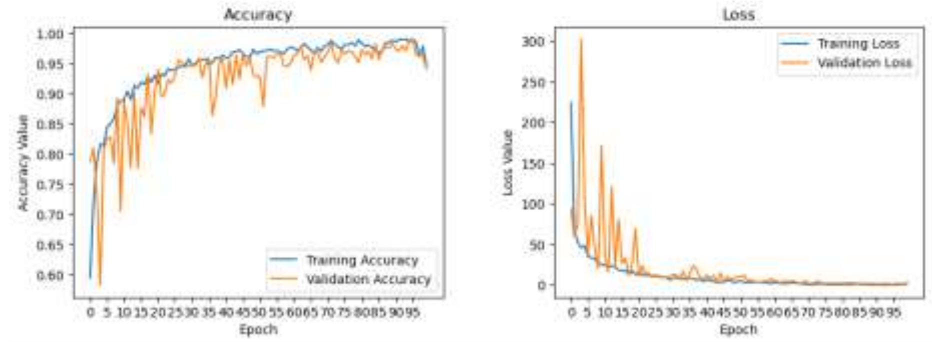
Fig 9. Comparison of multiclass classifiers with six optimizers based on accuracy(a) ANN model, (b) GRU model, and (c) TN model


Accuracy
Training Accuracy
Validation Accuracy
Accuracy Value
Epoch
Loss
Training Loss
Validation Loss
Loss Value
Epoch

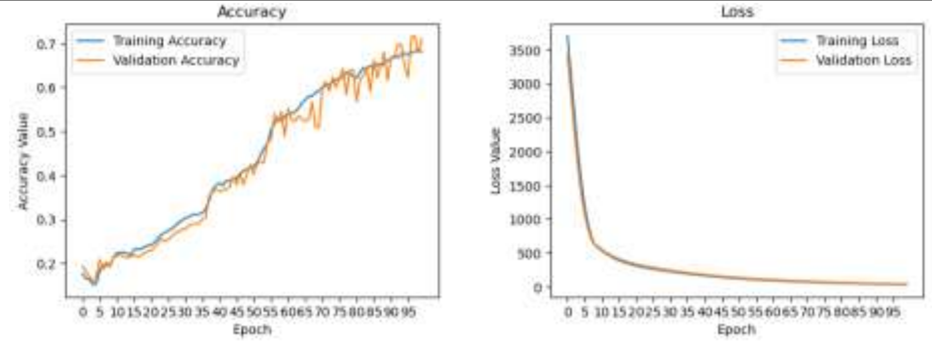
(a)


Accuracy
Training Accuracy
Validation Accuracy
Accuracy Value
Epoch
Loss
Training Loss
Validation Loss
Loss Value
Epoch

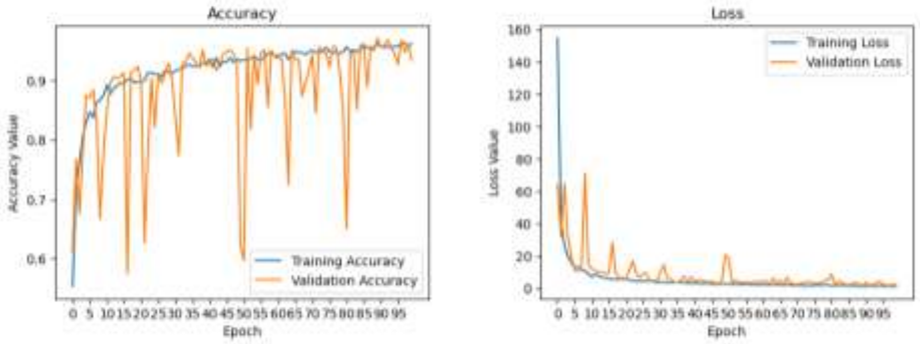
(b)


Accuracy
Training Accuracy
Validation Accuracy
Accuracy Value
Epoch
Loss
Training Loss
Validation Loss
Loss Value
Epoch

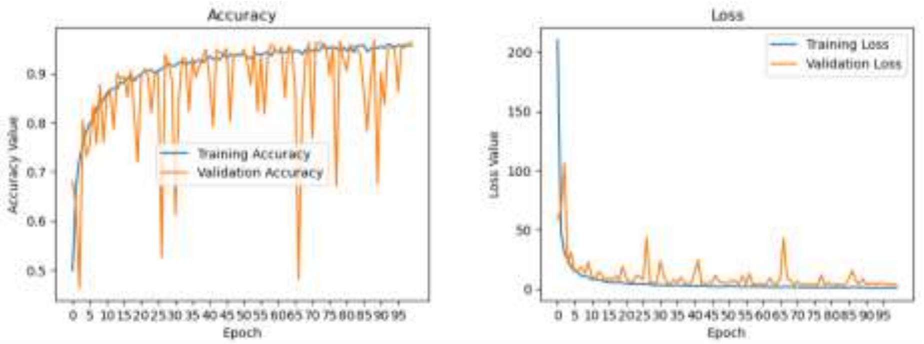
(c)


Accuracy
Training Accuracy
Validation Accuracy
Accuracy Value
Epoch
Loss
Training Loss
Validation Loss
Loss Value
Epoch

(d)

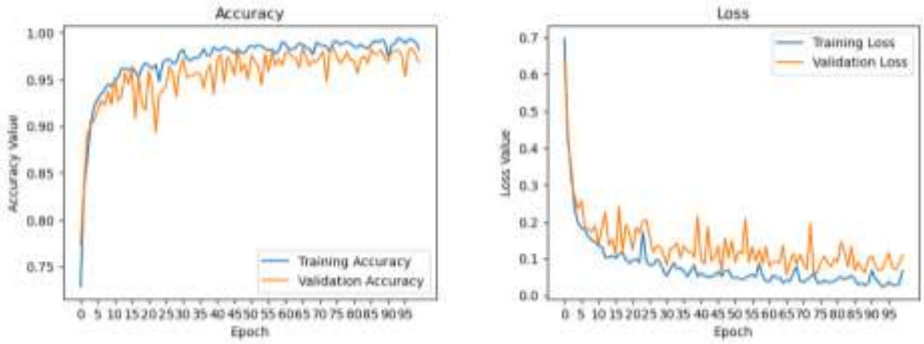


(e)

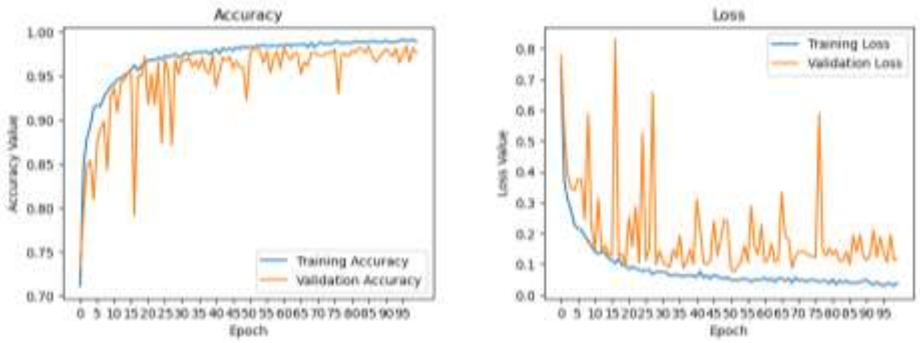


(f)

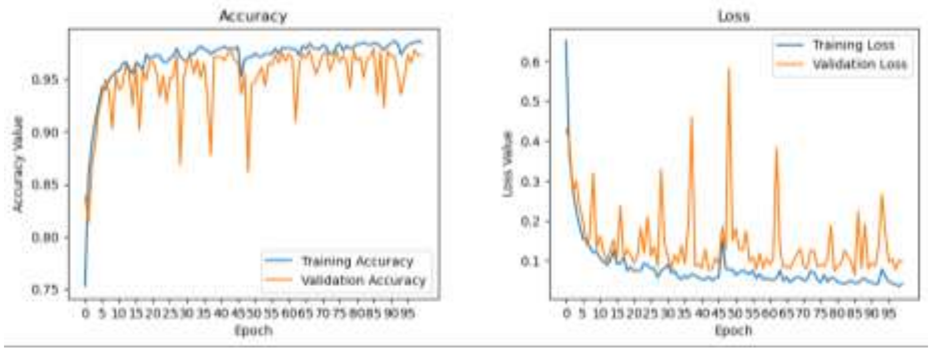
Fig 10. ANN accuracy performance curve with six optimizers (a) Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



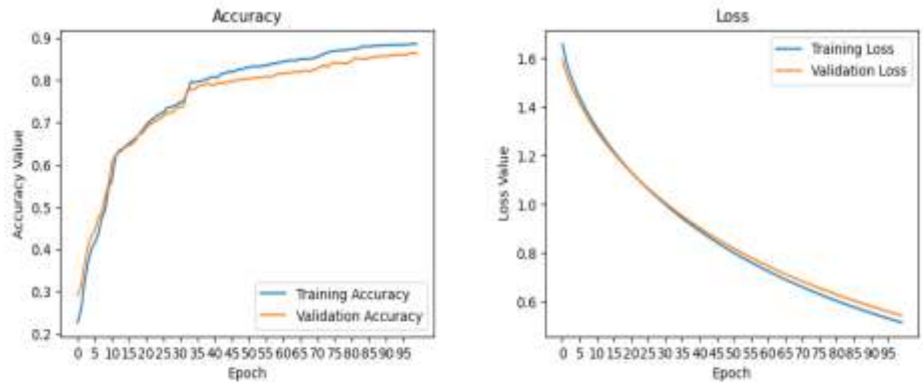
(a)



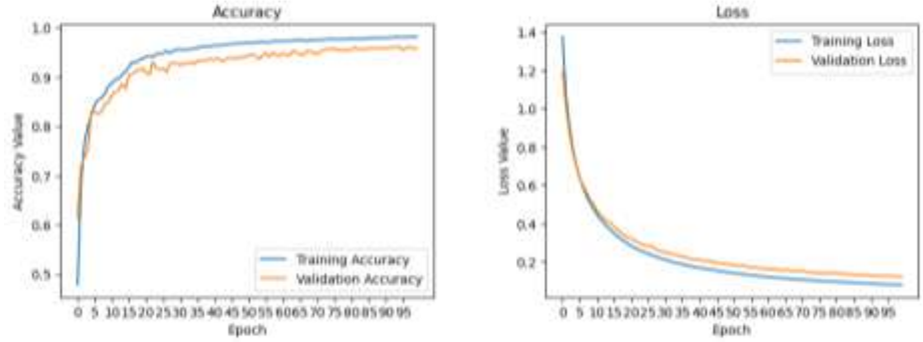
(b)



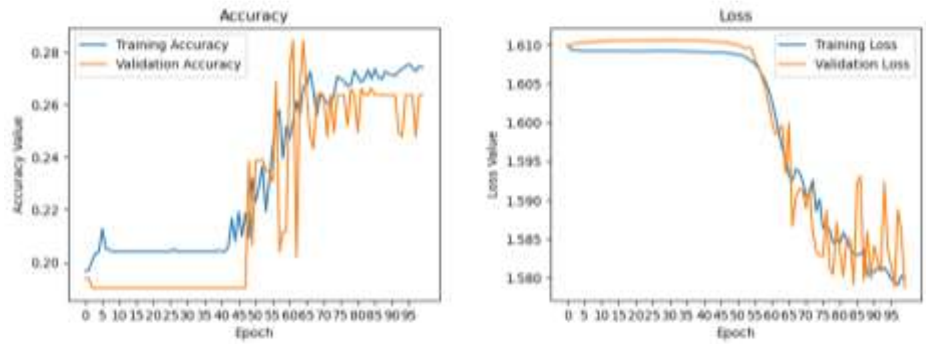
(c)



(d)

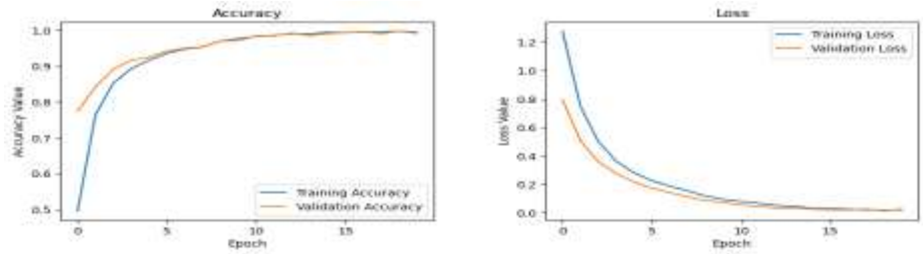


(e)

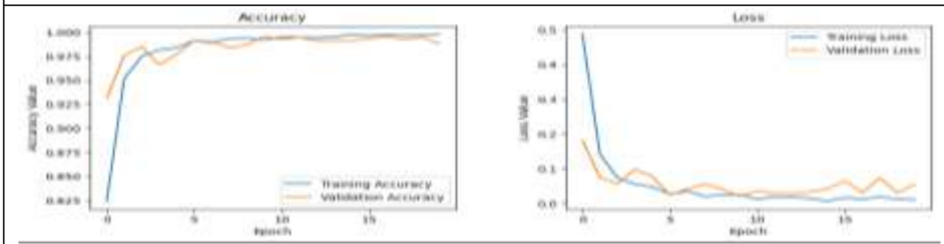


(f)

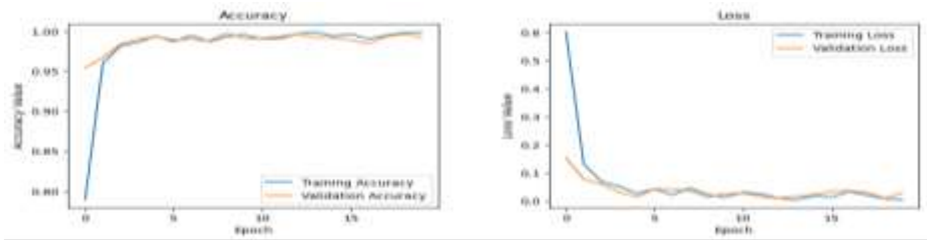
Fig 11. GRU accuracy performance curve with six optimizers (a) Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



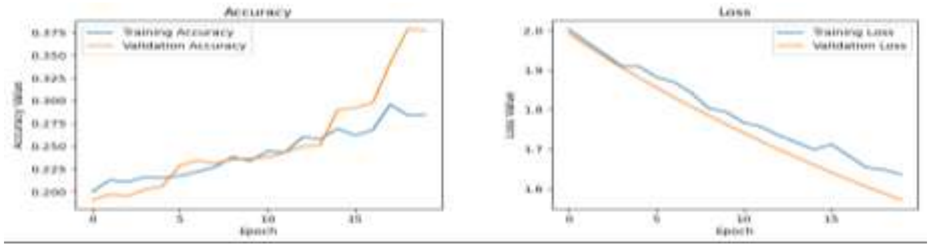
(a)



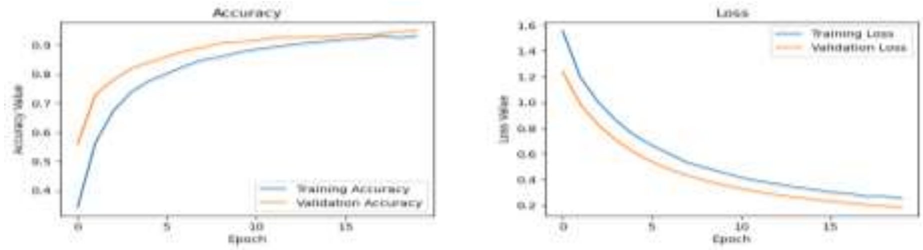
(b)



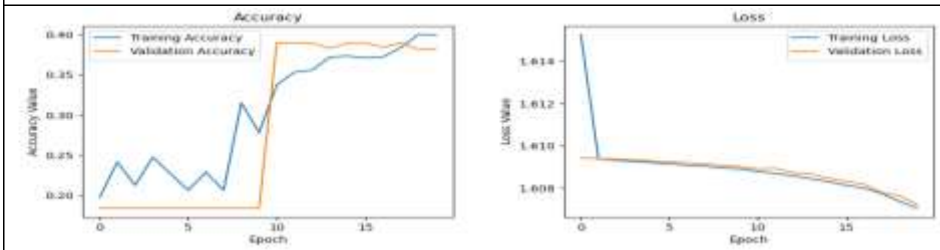
(c)



(d)

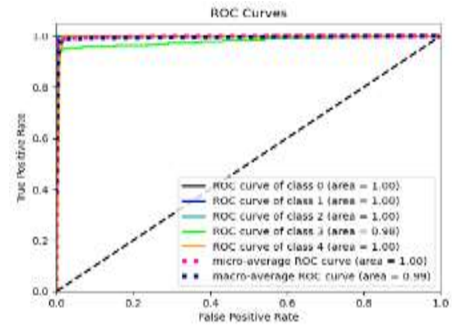


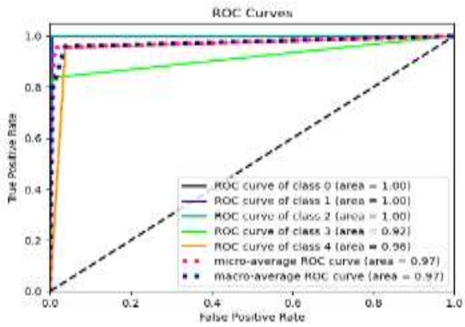
(e)



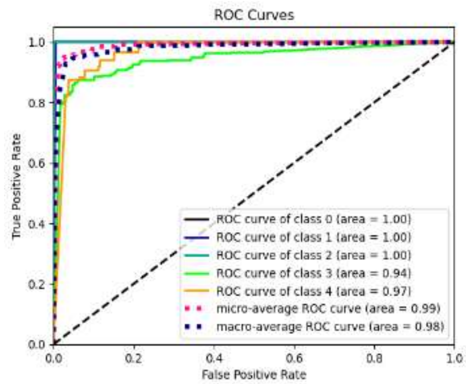
(f)

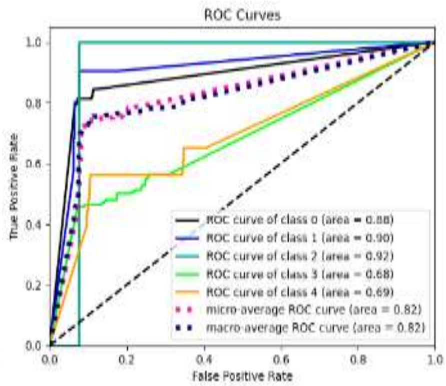
Fig 12. TN accuracy performance curve with six optimizers (a) Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



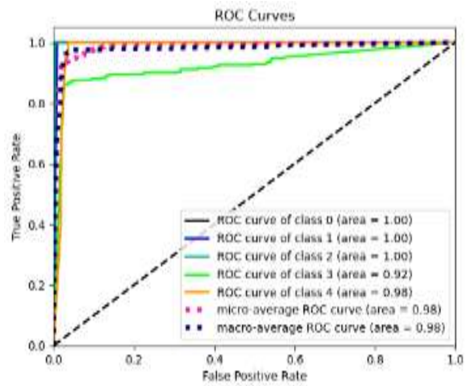


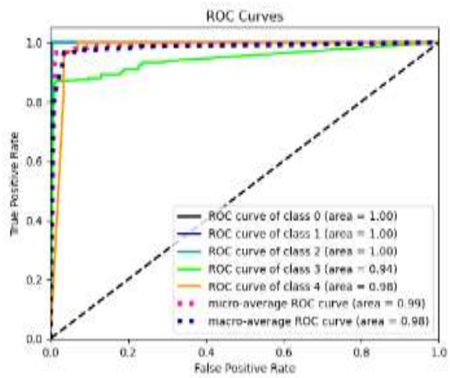
(a) **(b)**





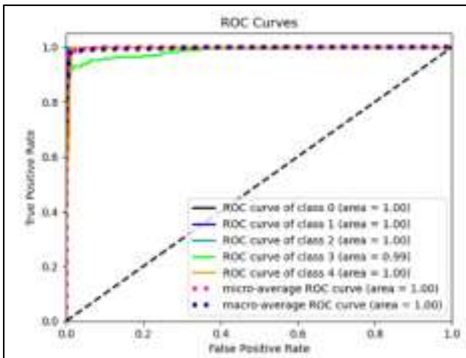
(c) **(d)**

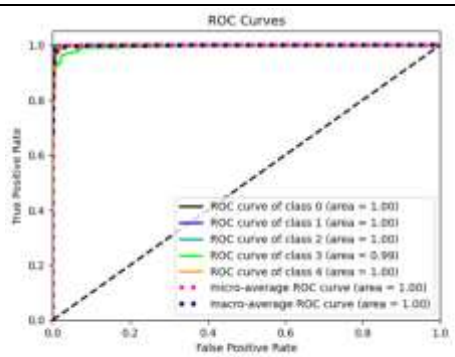




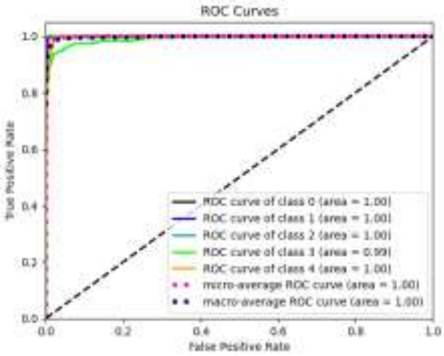
(e) **(f)**

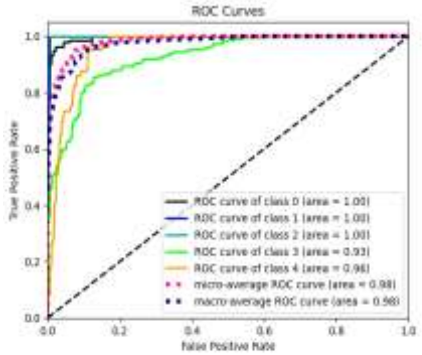
Fig 13. ANN performance with ROC-AUC on six optimizers (a) Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



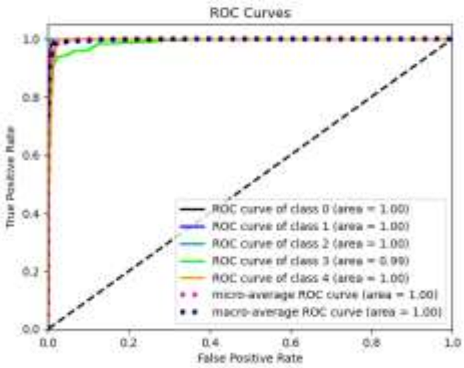


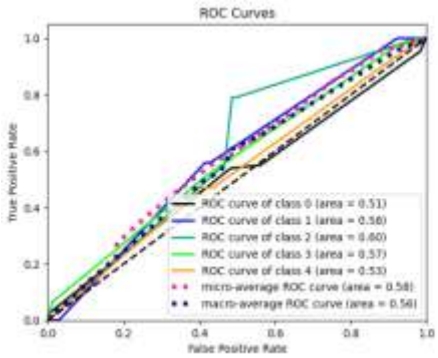
(a) **(b)**





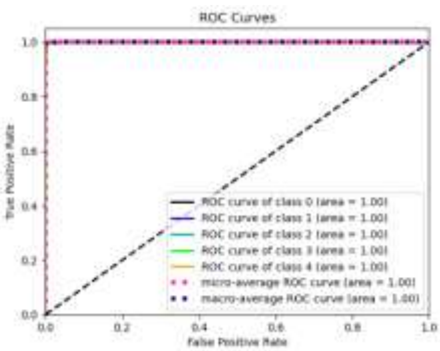
(c) **(d)**

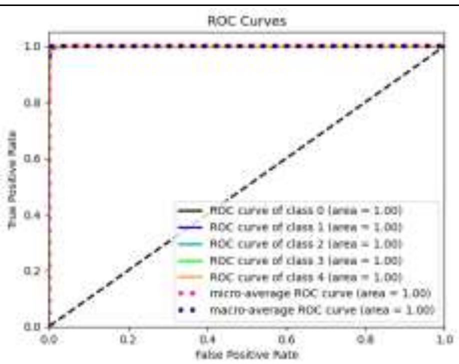




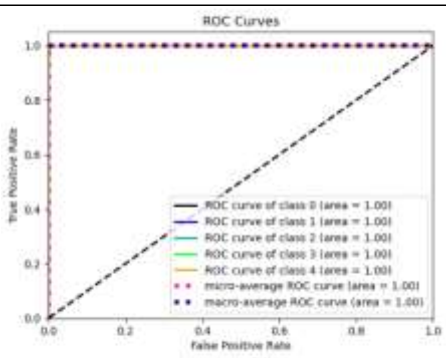
(e) **(f)**

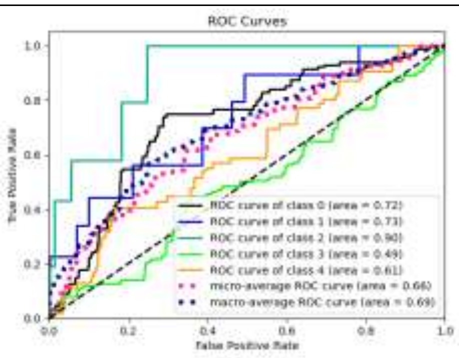
Fig 14. GRU performance with ROC-AUC on six optimizers (a) Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



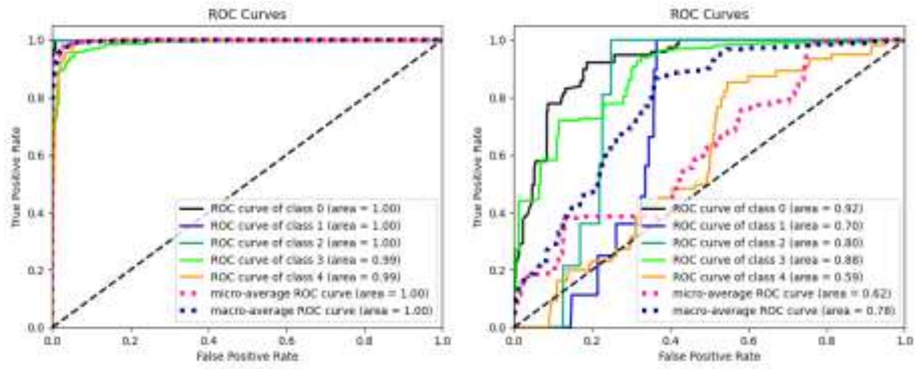


(a) **(b)**



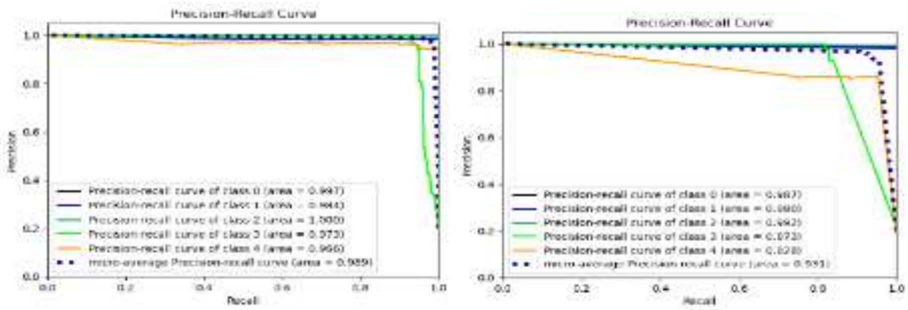


(c) **(d)**

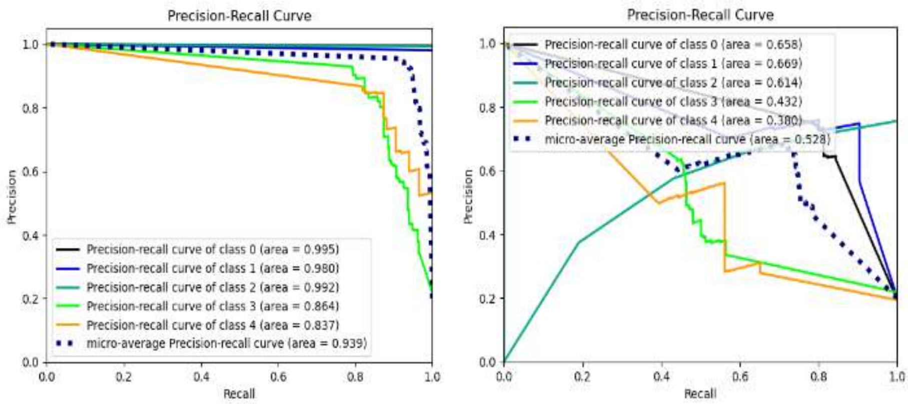


(e) (f)

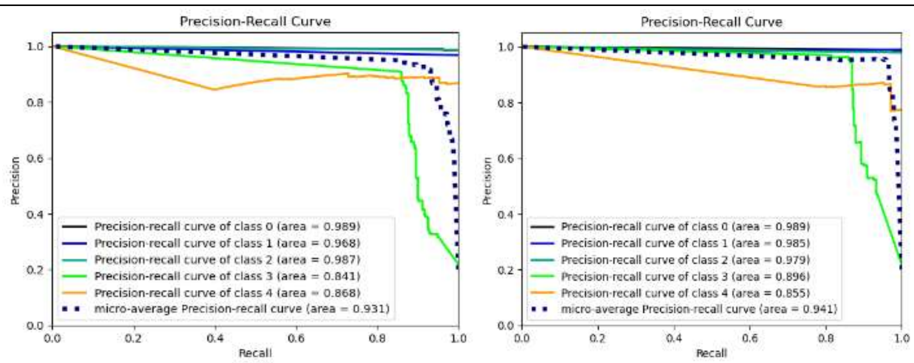
Fig 15. TN performance with ROC-AUC on six optimizers (a) Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



(a) (b)

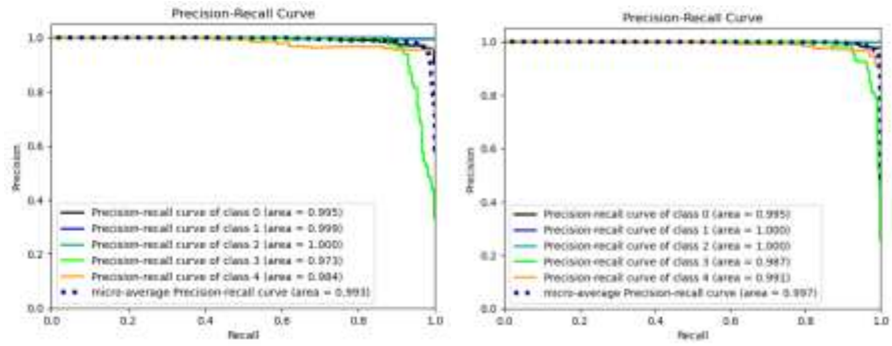


(c) (d)

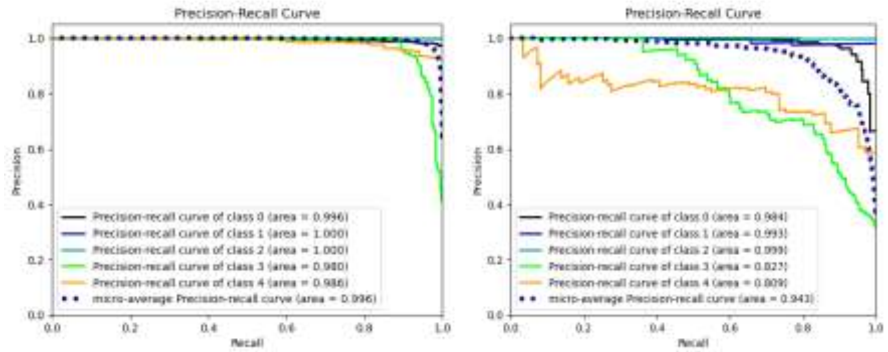


(e) (f)

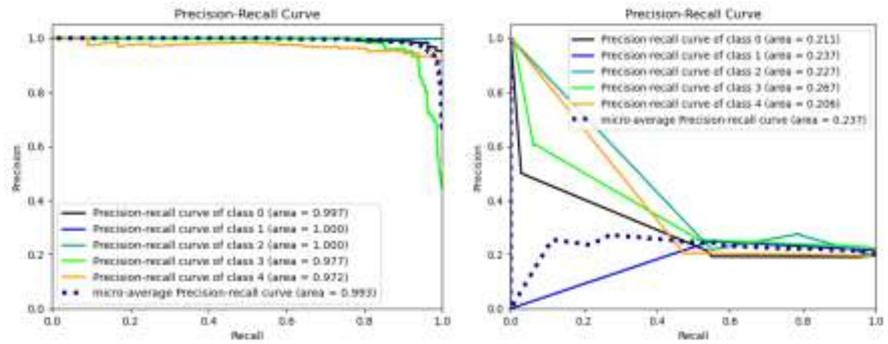
Fig 16. ANN performance with Precision-Recall Curve on six optimizers (a) Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



(a) (b)

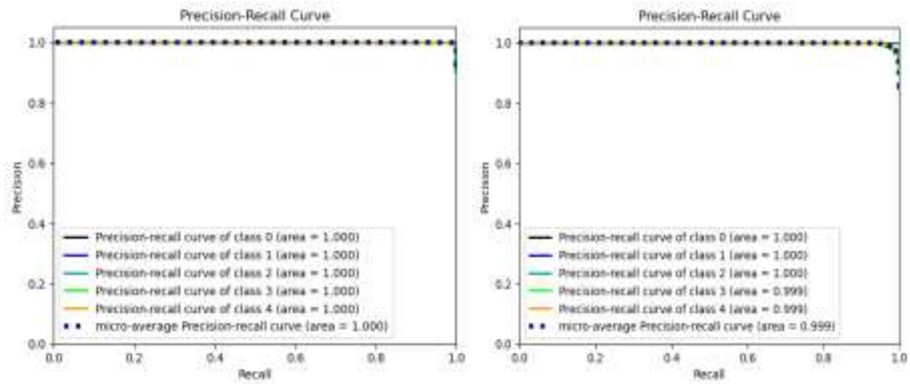


(c) (d)

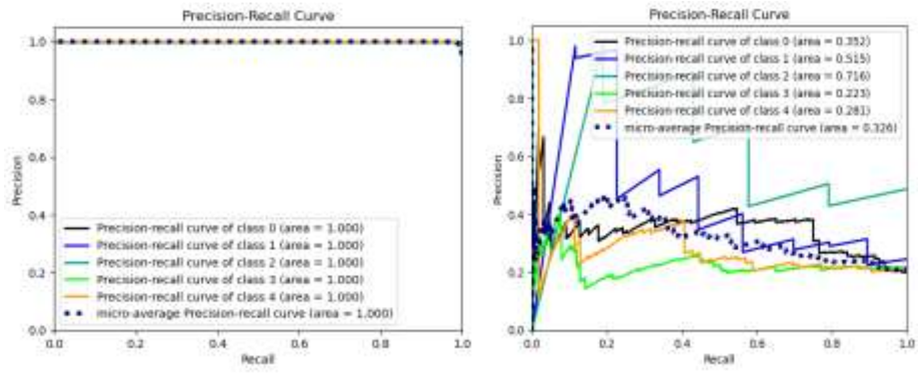


(e) (f)

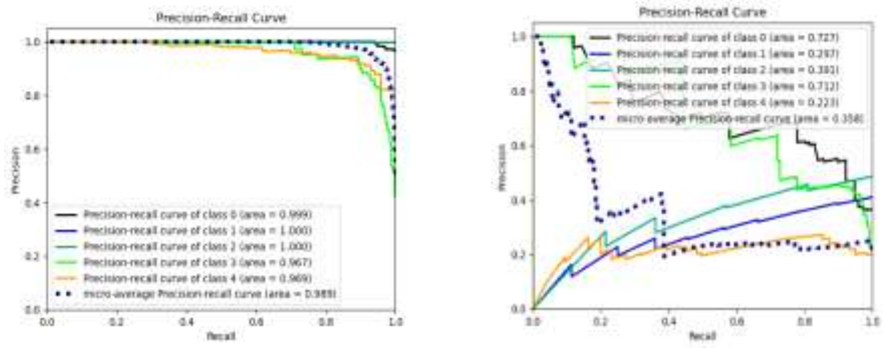
Fig 17. GRU performance with Precision-Recall Curve on six optimizers (a) Adam, b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl



(a) (b)



(c) (d)



(e) (f)

Fig 18. TN performance with Precision-Recall Curve on six optimizers (a)Adam, (b) RMSprop, (c) Nadam, (d) Adadelta, (e) Adagrad, (f) Ftrl

4 Result Analysis

The accuracy of the proposed classification models Artificial Neural Network (ANN), Gated Recurrent Unit (GRU), and Transformer Network (TN) are adequately evaluated using the metrics, roc-auc and precision-recall. Three models shows different metrics for the six optimizers for the evaluation of the models. Table 1-3. presents the accuracy measures of ANN, GRU, and TN across the six optimizers Adam, RMSprops, Nadam, Adadelta, Adagrad, and Ftrl. The bar chart in Fig 9. depicted the comparison of multiclass classifiers ANN, GRU, and TN with six optimizers based on accuracy. The above-mentioned comparison concludes that, the three optimizers Adam, RMSprops, and Nadam excels in all models than others. Notably, Adadelta and Ftrl failed with the accuracy below 0.70 in two models.

Even though the study also plots the roc-auc curve [38] for the multiclass classification models, showing the model's ability to differentiate among five classes. Fig. 10-12, depicts the accuracy performance curve with six optimizers of ANN, GRU and TN models. Fig. 13-15, shows the roc-auc performance curve for each models with six optimizers, providing better insights in the models performance. Fig. 16-18, plots the performance with Precision-Recall Curve [39] on six optimizers. Overall, the study highlights that, considering all the metrics the transformer network model with Adam and Nadam performed well with accuracy metrics 0.99. In addition, the study concludes that the advanced deep learning models shows better performance in multiclass classification with adequate metrics.

5 Conclusions

Recent progress in deep learning architectures have led to a significant boost in customer segmentation competence in the banking industry. By introducing the Artificial Neural Networks (ANN), Gated Recurrent Units (GRU), and Transformer Network (TN) models for the classification, businesses can acquire strong insights into the customer information, and identify bank ambassadors for better decision making. Thus, developing a business intelligent system can significantly boost the revenue. The proposed work discussed that while all three DL models perform well beyond all performance indicators, TN outperformed with its all capability to capture the intricate data patterns, hence, offering better classification ability. Future research can enhanced the study on hybrid deep learning approach and advanced optimized techniques utilized for classification.

References

1. Md Abu Sufian Mozumder, Fuad Mahmud, Md Shujan Shak, Nasrin Sultana, Gourab Nicholas Rodrigues, Md Al Rafi, Md Zahidur Rahman Farazi, Md Razaul Karim, Md. Sayham Khan, & Md Shahriar Mahmud Bhuiyan. (2024). Optimizing Customer Segmentation in the Banking Sector: A Comparative Analysis of Machine Learning Algorithms. Journal of Computer Science and Technology Studies, 6(4), 01–07. https://doi.org/10.32996/jcsts.2024.6.4.1.
2. N. Gankidi et al. "Customer Segmentation Using Machine Learning". In: (June 2022), pp. 1–5. DOI: 10.1109/CONIT55038.2022.9848389.
3. Dicuonzo, G., Galeone, G., Zappimbulso, E. and Dell'Atti, V., 2019. Risk management 4.0: The role of big data analytics in the bank sector. International Journal of Economics and Financial Issues, 9(6), pp.40-47.
4. More, R. and Moily, Y., 2021. Big data analysis in banking sector. International Journal of Engineering Research and Applications, 11(4), pp.1-5. ISSN: 2248-9622, DOI: 10.9790/9622-1104020105.
5. Utkarsh Srivastava, Santosh Gopalkrishnan, Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks, Procedia Computer Science, Volume 50, 2015, Pages 643-652, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.04.098.
6. T. Gupta, N. Gupta, A. Agrawal, A. Agrawal and K. Kansal, "Role of Big Data Analytics In Banking," 2019 International Conference on contemporary Computing and Informatics (IC3I), Singapore, 2019, pp. 222-227, doi: 10.1109/IC3I46837.2019.9055616.
7. Bansal, G. Katoch, N. Arora, A. Sharma, R. C. Bhadula and S. Agarwal, "Big data analytics in the Indian banking sector: An empirical study," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 1624-1627, doi: 10.1109/ICACITE53722.2022.9823738.
8. Milojević, Nenad and Redzepagic, Srdjan. "Prospects of Artificial Intelligence and Machine Learning Application in Banking Risk Management" Journal of Central Banking Theory and Practice, vol.10, no.3, 2021, pp.41-57. https://doi.org/10.2478/jcbtp-2021-0023.
9. Hassani, H., Huang, X., Silva, E. et al. Deep Learning and Implementations in Banking. Ann. Data. Sci. Volume 7, pages 433–446, (2020), https://doi.org/10.1007/s40745-020-00300-1.
10. Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T. and Ali, F., 2023. An advanced deep learning models-based plant disease detection: A review of recent research. Frontiers in Plant Science, 14, p.1158933.
11. Castelli, M., Vanneschi, L. and Largo, Á.R., 2018. Supervised learning: classification. por Ranganathan, S., M. Grisbskov, K. Nakai y C. Schönbach, 1, pp.342-349.
12. Kotsiantis, S.B., Zaharakis, I. and Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), pp.3-24.
13. Sen, P.C., Hajra, M. and Ghosh, M., 2020. Supervised classification algorithms in machine learning: A survey and review. In Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018 (pp. 99-111). Springer Singapore.
14. P. Hemalatha and G. M. Amalanathan, "A Hybrid Classification Approach for Customer Churn Prediction using Supervised Learning Methods: Banking Sector," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019, pp. 1-6, doi: 10.1109/ViTECoN.2019.8899692.
15. Dalmia, H., Nikil, C.V.S.S., Kumar, S. (2020). Churning of Bank Customers Using Supervised Learning. In: Saini, H.S., Singh, R.K., Tariq Beg, M., Sahambi, J.S. (eds) Innovations in Electronics and Communication Engineering. Lecture Notes in Networks and Systems, vol 107. Springer, Singapore. https://doi.org/10.1007/978-981-15-3172-9_64.
16. J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, Nigeria, 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782.
17. Alex G.C. de Sá, Adriano C.M. Pereira, Gisele L. Pappa, A customized classification algorithm for credit card fraud detection, Engineering Applications of Artificial Intelligence, Volume 72, 2018, Pages 21-29, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2018.03.011.
18. Bocklitz, T., Putsche, M., Stüber, C., Käs, J., Niendorf, A., Rösch, P. and Popp, J., 2009. A comprehensive study of classification methods for medical diagnosis. Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering, 40(12), pp.1759-1765.

19. Inbarani, H.H., 2015. A novel neighborhood rough set based classification approach for medical diagnosis. Procedia Computer Science, 47, pp.351-359.
20. Thein, H.T.T. and Tun, K.M.M., 2015. An approach for breast cancer diagnosis classification using neural network. Advanced Computing, 6(1), p.1.
21. Nadjate Saidani, Kamel Adi, Mohand Saïd Allili, A semantic-based classification approach for an enhanced spam detection, Computers & Security, Volume 94, 2020, 101716, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2020.101716.
22. P. S. Parmar, P. K. Biju, M. Shankar and N. Kadiresan, "Multiclass Text Classification and Analytics for Improving Customer Support Response through different Classifiers," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018, pp. 538-542, doi: 10.1109/ICACCI.2018.8554881.
23. Anita Prinzie, Dirk Van den Poel, Random Forests for multiclass classification: Random MultiNomial Logit, Expert Systems with Applications, Volume 34, Issue 3, 2008, Pages 1721-1732, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2007.01.029.
24. Sonule, Avinash & Mendhe, Vikas. (2023). Customer Segmentation Using Artificial Neural Network. 2320-2882, International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org.
25. Derrick S. Boone, Michelle Roehm, Retail segmentation using artificial neural networks, International Journal of Research in Marketing, Volume 19, Issue 3, 2002, Pages 287-301, ISSN 0167-8116, https://doi.org/10.1016/S0167-8116(02)00080-0.
26. Vellido, P.J.G. Lisboa, K. Meehan, Segmentation of the on-line shopping market using neural networks, Expert Systems with Applications, Volume 17, Issue 4, 1999, Pages 303-314, ISSN 0957-4174, https://doi.org/10.1016/S0957-4174(99)00042-1.
27. Ozan, Ş. and Iheme, L.O., 2019, April. Artificial neural networks in customer segmentation. In 2019 27th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.
28. Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. and Arshad, H., 2018. State-of-the-art in artificial neural network applications: A survey. Heliyon, 4(11).
29. https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/
30. https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa.
31. Elsayed, N., Maida, A.S. and Bayoumi, M., 2019, July. Gated recurrent neural networks empirical utilization for time series classification. In 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (pp. 1207-1210). IEEE.
32. https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2.
33. https://www.analyticsvidhya.com/blog/2021/03/introduction-to-gated-recurrent-unit-gru/
34. Cheng, X., Lin, H., Wu, X., Shen, D., Yang, F., Liu, H. and Shi, N., 2022, July. Mltr: Multi-label classification with transformer. In 2022 IEEE international conference on multimedia and expo (ICME) (pp. 1-6). IEEE.
35. https://exchange.scale.com/public/blogs/how-to-build-a-transformer-for-supervised-classification.
36. https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture).
37. Negash, S. and Gray, P., 2008. Business intelligence. Handbook on decision support systems 2, pp.175-193.
38. Z. Yang, Q. Xu, S. Bao, X. Cao and Q. Huang, "Learning with Multiclass AUC: Theory and Algorithms," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7747-7763, 1 Nov. 2022, doi: 10.1109/TPAMI.2021.3101125.
39. Tharwat, A. (2021), "Classification assessment methods", Applied Computing and Informatics, Vol. 17 No. 1, pp. 168-192. https://doi.org/10.1016/j.aci.2018.08.003.