iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73547

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Enhanced Customer Retention: Deep Learning-Based Churn Prediction for Telecom Industry

S Surekha[1], B Yamuna[2]
[1]Assistant Professor, Department of CSE, UCEK (A) JNTU Kakinada, Andhra Pradesh
[2]Student, Master of Computer Applications, UCEK (A) JNTU Kakinada, Andhra Pradesh

Abstract: The Telecommunications Industry (TCI) faces a significant problem with customer attrition since revenue generation depends on keeping current customers. A deep learning-based architecture makes more accurate predictions about customer attrition. This brings to light a lot of problems. Churn aids in prioritizing new features or services that have the best chance of increasing customer retention. This guarantees that resources are directed toward the areas that can reduce churn the most.In this research, a deep learning-based model for predicting customer attrition in the telecom sector is presented. In order to extract sequential patterns from consumer behaviour, the model uses a 1D Convolutional Neural Network (CNN). By identifying spatial links in customer data, a 2D CNN improves feature extraction.The telecom statistics available on the Kaggle website aid in the prediction of churn in the telecom sector. Class imbalance in datasets is addressed by using the SMOTE, SMOTEEN, and SMOTETomek approaches. The performance analysis assesses recall rate, accuracy, precision, and F1 score. This methodical approach improves prediction accuracy.
Keywords: Churn prediction, Deep learning, CNN, FNN, SMOTETomek, SMOTEEN, SMOTE.

I. INTRODUCTION

The telecom sector has expanded significantly over time, and consumers now have a wide range of options and service providers to pick from. As a result, competition is now fierce. Customers can now more easily switch companies if they find better deals or services. Because of this, telecom companies are facing a significant issue with customer churn, which occurs when a customer discontinues using their service. Large financial losses can result from even a slight increase in the number of customers quitting [1]. Churn directly affects a business's earnings. It is well known that attracting new clients is far more costly than keeping current ones [2].In addition to losing revenue, firms that lose consumers often have to spend more money to find new ones. By keeping clients for longer, telecom operators can boost lifetime value, foster trust, and promote customer loyalty. As a result, a major corporate objective for sustained success is to comprehend and avoid churn [3].

To anticipate turnover early, telecom companies and other researchers have begun utilising customer data. Before a consumer decides to leave, this enables businesses to take preventive measures like enhancing customer service or providing tailored offers. Proactively detecting and resolving customer unhappiness has been demonstrated to greatly increase retention [4], [5]. Through the examination of consumer behaviour, service utilisation, and interaction history, companies can make more informed decisions to lower customer attrition and maintain customer satisfaction [6].

However, because churn behaviour incorporates a variety of elements, including user preferences, service quality, and outside influences, it is complicated and frequently har d to anticipate. In order to effectively reduce churn, traditional approaches have frequently failed to adequately capture these processes [7]. Recent research indicates that innovative strategies are working better, particularly those that can better identify important aspects and comprehend client patterns [8], [9]. With these better tactics, telecom companies are avoiding churn by anticipating and mitigating it before it occurs, which strengthens customer connections and gives them a competitive edge in the market.

II. RELATED WORK

A deep learning model that was optimised with optimisation algorithms was presented by Nagaraju et al. [1] using stacked bidirectional LSTM and RNN. It uses more efficient data processing to anticipate health insurance client attrition with high accuracy.

To find churn customers in the telecom industry, Irfan Ullah et al. [2] employed Random Forest and additional machine learning techniques. They achieved an accuracy of 88.63% by classifying customers according to their similarities and recommending retention tactics.

Somak Saha and colleagues et al. [3] used spatial attention, residual blocks, and 1D CNN to create the deep learning model ChurnNet. Using SMOTE-based techniques, it addressed class imbalance and, across several datasets, achieved over 95% accuracy.
Syed Fakhar Bilal et al. [4] presented a hybrid churn prediction model that blends clustering (like K-means) with classification methods. The ensemble model functioned well on telecom data, obtaining 94.7% accuracy.
A BiLSTM-CNN model was developed by Khattak et al. [5] to enhance churn prediction. It demonstrated improved accuracy on benchmark datasets and addressed problems such as inadequate feature encoding.
In order to correct for data imbalance in churn prediction, Adnan Amin et al. [6] compared oversampling methods like SMOTE and MTDF. They discovered that the best outcomes for enhancing model performance came from MTDF and genetic algorithms. An attention-based BiLSTM-CNN model for financial churn prediction was introduced by Y. Liu et al. [7]. By assisting the model in concentrating on key elements, the attention mechanism marginally increased accuracy and dependability.
FE Using decision forest models, such as random forest and logistic model trees, was suggested by Usman-Hamza et al. [8]. These performed better than previous machine learning techniques and handled unbalanced telecom data well.
IV Pustokhina et al. [9] presented a model that combines an optimised learning algorithm (OWELM) with enhanced SMOTE. It achieved good results on telecom datasets by fine-tuning parameters using the Rain Optimisation Algorithm.
A genetic algorithm combined with Naïve Bayes was used by Amin et al. [10] to create an adaptive churn prediction model. It achieved up to 98% accuracy on several telecom datasets and adapts to changing user behaviour.

III. METHODOLOGY

A churn prediction model based on deep learning is shown in this study employing telecom. The churn data dataset includes 3,333 customer details. 483 of the total consumers left, while 2,850 remained. Figure 1: The three primary stages of the suggested system's methodology are preprocessing, model training, and testing.

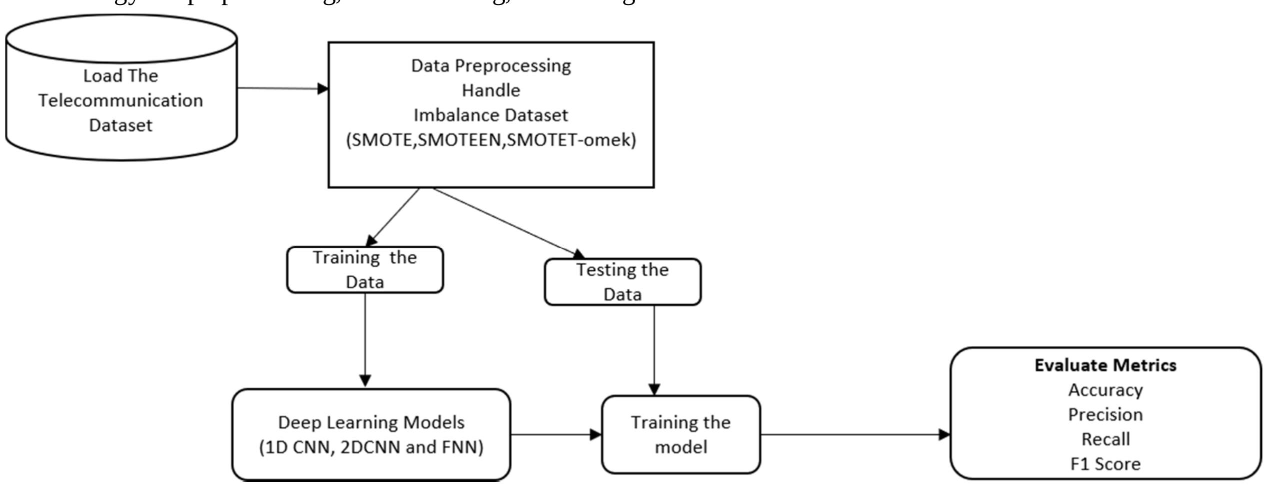


Figure 1: workflow of the proposed system

A. Data Preprocessing

Churn prediction begins with preprocessing. We handled missing values, normalised features using StandardScaler, turned categorical "yes/no" values to 1s and 0s, and cleaned the data by eliminating unnecessary columns. We used SMOTE, SMOTEENN, and SMOTETomek to improve model accuracy and correct class imbalance.

1) SMOTE: SMOTE generates points between adjacent similar data points to produce new, fictitious instances of the minority class (e.g., churned customers). Increasing the number of minority samples helps balance the dataset.
2) *SMOTETomek:* SMOTE Tomek detects and eliminates borderline samples that are near another class by combining SMOTE with Tomek Links. This improves model performance by cleaning up overlapping data points and balancing the dataset. These methods are essential for raising the churn prediction models' fairness and accuracy.
3) *SMOTEENN:* SMOTE adds fresh minority class samples, while ENN eliminates noisy or incorrectly categorised samples from both classes. This combination is called SMOTEENN. Because of this, the dataset is cleaner and more suitable for training precise models.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

B. Model Training
The cleaned and balanced dataset is used to train the deep learning models 1D CNN, 2D CNN, and FNN, which the system uses to identify trends and predict which clients are likely to leave.

1) 1D CNN
Time series, text data, and sensor signals are examples of sequential data that can be processed by a specific type of deep learning model known as a 1D CNN. Conventional CNNs analyse 2D images, whereas 1D CNNs process data with a sequential structure.1D CNN contain nine layers.

Input Layer

Receives a 1D sequence of reshaped customer data. It simply forwards the data to the following layer without processing it. Conv1D Layer

Employs 48 size 3 filters to find trends in the data, such as abrupt increases in usage. ReLU activation aids in emphasising significant signals.

MaxPooling1D Layer

Reduces noise and expedites processing by shrinking the data by retaining only the maximum values in windows of size 2.

Flatten Layer

prepares the 2D output from the preceding layer for dense (fully connected) layers by converting it into a 1D vector. Dense Layer 64 neurones make up this fully connected layer, which learns feature combinations to improve churn risk prediction. ReLU is used to account for non-linearity.

Dropout Layer

During training, half of the neurones are randomly turned off to avoid overfitting and aid in the model's generalisation.

Output Dense Layer

a single neurone with a churn probability ranging from 0 to 1 due to sigmoid activation. Higher churn risk is indicated by a value closer to 1.

2) 2D CNN
A specific type of deep learning model designed to analyse picture or grid-like input is called a 2D CNN. To extract important features from the input data, the model makes use of 2D convolutional filters, also known as kernels, in a 2D CNN. Each filter explores the input in both width and height dimensions in order to identify patterns.2D CNN contains six layers.

Input Layer

Customer data is reshaped and sent to the input layer in a square, two-dimensional, one-channel format. It merely gets the input ready for convolution without performing any calculations.

Conv2D Layer

learns spatial patterns between nearby features by applying 32 filters. ReLU aids in the identification of significant non- linear relationships. The output size remains constant with the same padding.

MaxPooling2D Layer

takes the maximum value from 2x2 blocks to reduce the size of the feature map. simplifies and preserves the most crucial elements.

Flatten Layer

Transforms the 2D data into a 1D vector so that dense layers can use it to make decisions.

Dense Layer

a fully connected layer that uses ReLU activation to discover deeper relationships from the patterns that were extracted.

Dropout Layer

60% of neurones are randomly turned off during training to enhance generalisation and avoid overfitting.

Output Layer
Produces a single number between 0 and 1 that indicates the likelihood of churn; a value nearer 1 indicates a higher likelihood of leaving.

3) FNN

ML using a certain type of artificial neural network. Information flows in a single direction through hidden layers from the input layer to the output layer.In the network, data simply flows ahead; there are no cycles or loops. FNN contain three layers.
Input Layer
obtains customer information such as plan types, charges, and call duration. It simply forwards the data to the following layer without processing it.

First Dense Layer
possesses 128 neurones and learns early patterns in the data, like the relationship between specific features and churn, through ReLU activation.

First Batch Normalization Layer
By normalising the output values from the dense layer, the first batch normalisation layer maintains their stability and expedites training.

First Dropout Layer
During training, 30% of the neurones are randomly turned off to avoid overfitting and enhance generalisation. Second Dense Layer
It keeps finding valuable patterns in the data with 64 neurones and ReLU activation. Second Batch Normalization Layer
To guarantee more seamless and quick learning, the second batch normalisation layer once more stabilises the data before moving on to the next layer.

Second Dropout Layer
Bpatterns.
Third Dense Layer consists of 32 neurones and aids in feature refinement, focusing on the most crucial details for the ultimate prediction. Output Layer
The probability of churn is output by a single neurone with sigmoid activation; a value closer to 1 indicates a higher likelihood of churn.

4) Model Testing

Performance Evaluation Metrics

The performance of a classification model can be assessed with the use of evaluation measures. They demonstrate how exact, accurate, and trustworthy the model's predictions are. Important metrics include F1 score, recall, accuracy, and precision.

a) Accuracy

The number of accurate predictions both positive and negative out of all the predictions is known as accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

b) Precision

Precision tells us how many of the samples that the model predicted as positive are actually positive.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

c) Recall

Recall is defined as the proportion of correctly detected true positives.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

d) F1-Score

The F1-score is the harmonic mean of recall and precision.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

IV. RESULTS AND ANALYSIS

The Table 1 shows the compares between 1D CNN, 2D CNN, and FNN without using the balancing techniques SMOTE, SMOTEENN, and SMOTETomek worked on Telecon Churn dataset based on several characteristics, such as accuracy, precision, recall, and F1-score.

Table 1: Performance Comparsion of Models

DL Models	Accuracy	Precision	Recall	F1-Score
FNN	0.886	0.585	0.851	0.694
2D CNN	0.874	0.559	0.802	0.659
1D CNN	0.816	0.435	0.723	0.543

The Table2 shows the compares between 1D CNN, 2D CNN, and FNN after using the balancing techniques SMOTE, SMOTEENN, and SMOTETomek worked on Telecon Churn dataset based on several characteristics, such as accuracy, precision, recall, and F1-score.

Table 2: Performance Comparsion of Models Applying Balancing Techniques.

DL Models	Data Balancing Techniques	Accuracy	Precision	Recall	F1-Score
FNN	SMOTE	0.941	0.936	0.944	0.940
	SMOTEENN	0.967	0.980	0.962	0.971
	SMOTETomek	0.925	0.924	0.923	0.923
2D CNN	SMOTE	0.891	0.883	0.895	0.889
	SMOTEENN	0.912	0.934	0.909	0.921
	SMOTETomek	0.875	0.872	0.872	0.872
1D CNN	SMOTE	0.811	0.869	0.719	0.787
	SMOTEENN	0.803	0.948	0.692	0.800
	SMOTETomek	0.811	0.887	0.703	0.784

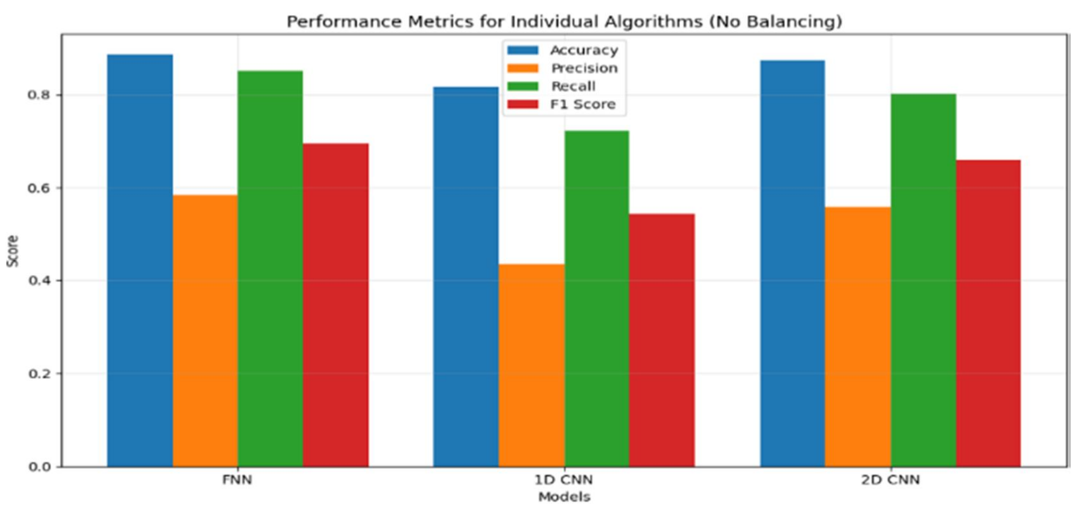


Figure 2: Comparison Graphs for 1D CNN, 2D CNN, and FNN

Figure 2 displays the performance of FNN, 1D CNN, and 2D CNN without data balancing. With the highest F1-score (0.694), accuracy (88.6%), and recall (85.1%), FNN was the most effective model. With slightly lower results, 2D CNN landed in second place, while 1D CNN did the worst across the board.

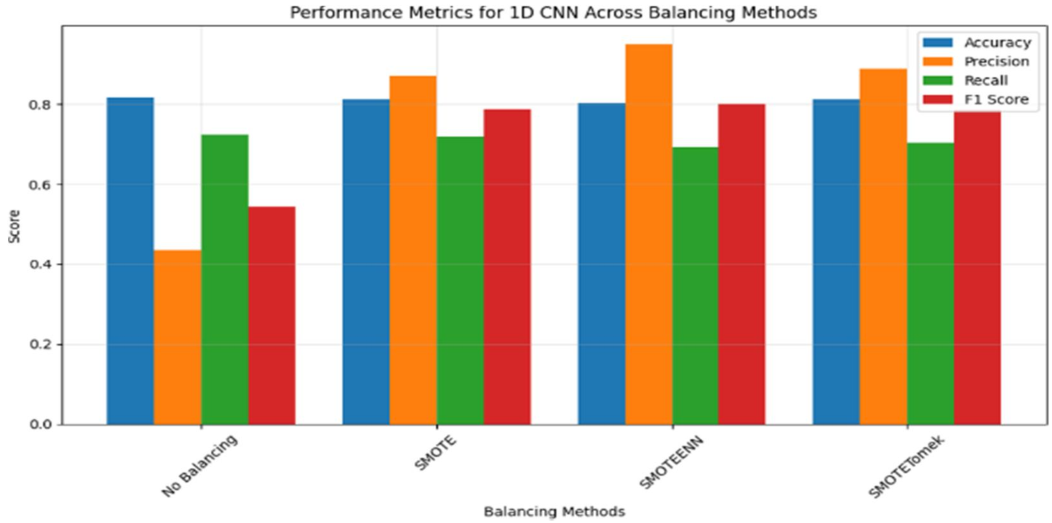


Figure 3: Comparison Graph for 1D CNN Applying Balancing Techniques

The 1D CNN model performs better when all three balancing strategies are used, as shown in Figure 3: While SMOTEENN gets the highest precision (94.8%) and F1-score (0.800), SMOTE delivers a decent balance of metrics, while SMOTETomek yields reliable and consistent results. While all approaches enhance the model in general, SMOTEENN outperforms them all.

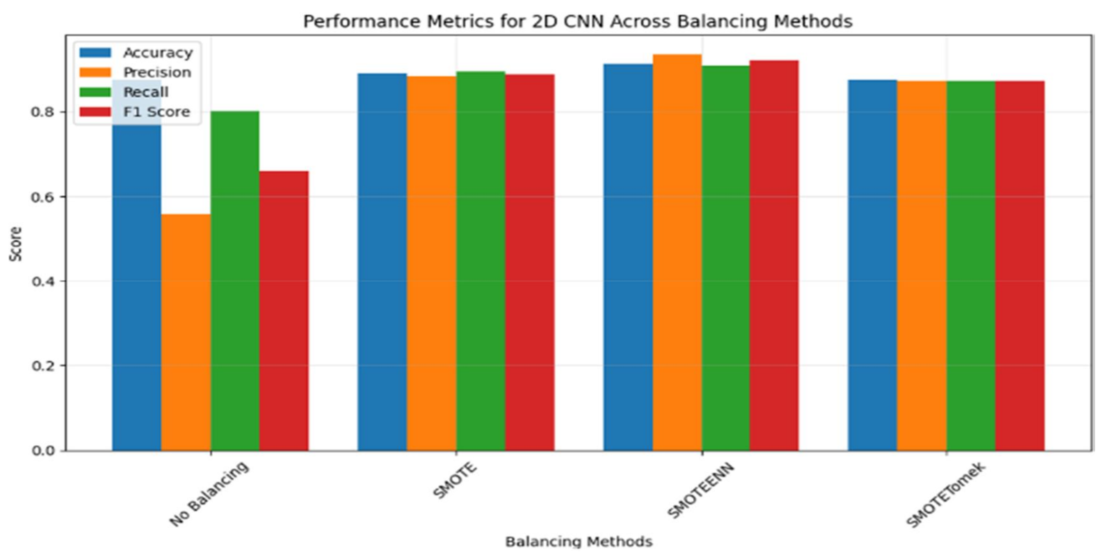


Figure 4: Comparison Graph for 2D CNN Applying Balancing Techniques.

Figure 4 shows The application of balancing strategies significantly improves the performance of the 2D CNN model. All metrics, including accuracy, precision, recall, and F1-score, reach about 89% with SMOTE, while SMOTEENN performs best, with precision slightly above 93% and recall and F1-score close to 91%. This demonstrates strong and well-rounded performance. Furthermore, SMOTETomek produces reliable results, with all metrics averaging 87.5%. When everything is said and done, balancing strategies improve the accuracy of the model, with SMOTEENN doing better than the others.

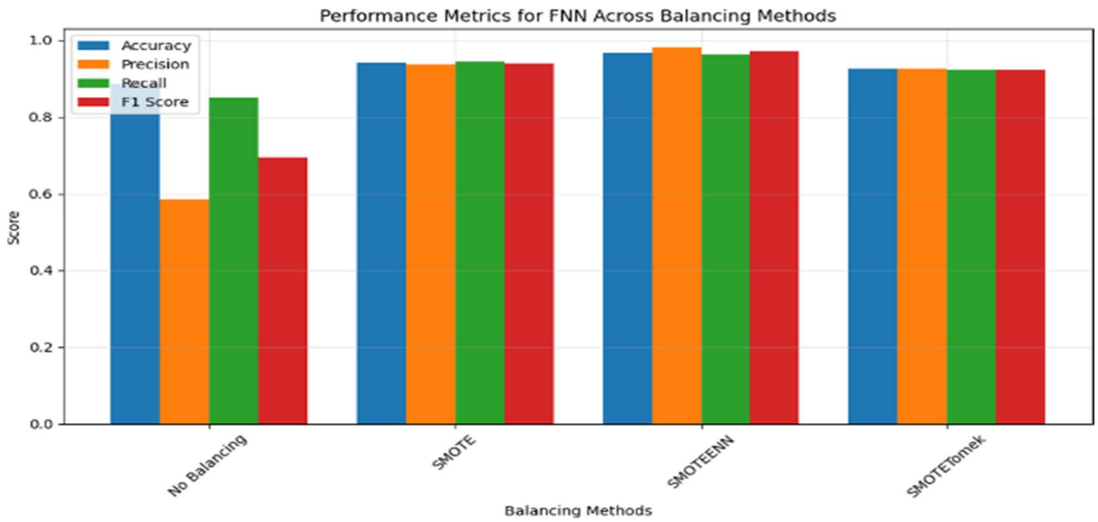


Figure 5: Comparison Graph for FNN Applying Balancing Techniques.

Figure 5 illustrates the FNN model's performance using a variety of balancing strategies. SMOTE's predictions are reliable and well-balanced; its accuracy, precision, recall, and F1-score all approach 94%. With a 97.1% F1-score, 96.7% accuracy, 98% precision, and 96.2% recall, SMOTEENN yields the best results. Furthermore, all indicators for SMOTETomek hover around 92.5%, indicating good performance. Generally speaking, the FNN model performs better than deep learning models.

A. Discussion

The paper's findings clearly demonstrate that using data balancing strategies such as SMOTE, SMOTEENN, and SMOTETomek significantly improves churn prediction. The FNN, 1D CNN, and 2D CNN deep learning models were the three that were tested. The best-performing model with the highest accuracy was the FNN model using SMOTEENN. This demonstrates how basic models can produce positive results when combined with appropriate data preparation. However, the FNN model's performance was still superior to that of the 1D CNN and 2D CNN models, even though they both improved when balancing techniques were used. Overall, a dependable system to more precisely forecast customer attrition and lower losses in the telecom sector may be built by utilising the appropriate model and data handling technique.

V. CONCLUSION

The purpose of this paper was to use deep learning to build a reliable system for predicting customer churn in the telecom sector. To solve the class imbalance problem, three models FNN, 1D CNN, and 2D CNN were tested along with balancing techniques like SMOTE, SMOTETomek, and SMOTEENN. Each model was evaluated using accuracy, precision, recall, F1-score, and confusion matrices. Among them, the FNN model combined with SMOTEENN gave the best results, with accuracy improving from 88.6% to 96% and the F1-score increasing from 64 to 94. This shows that even a simple but well-tuned model can outperform more complex ones when working with structured data.

REFERENCES

[1] N. Jajam, N. P. Challa, K. S. L. Prasanna, and C. H. V. S. Deepthi, “Arithmetic Optimisation With Ensemble Deep Learning SBLSTM-RNN-IGSA Model for Customer Churn Prediction,” IEEE Access, vol. 11, pp. 93111–93128, 2023, doi: 10.1109/ACCESS.2023.3304669.

[2] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, “A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector,” IEEE Access, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[3] S. F. Bilal, A. A. Almazroi, S. Bashir, F. H. Khan, and A. A. Almazroi, “An ensemble-based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in the telecom industry,” PeerJ Comput. Sci., vol. 8, p. e854, Feb. 2022.

[4] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, “Customer churn prediction using composite deep learning technique,” Sci. Rep., vol. 13, no. 1, Oct. 2023, Art. no. 17294.

[5] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, and A. Hussain, “Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study,” IEEE Access, vol. 4, pp. 7940–7957, 2016.

[6] Y. Liu, M. Shengdong, G. Jijian, and N. Nedjah, “Intelligent Prediction of Customer Churn with a Fused Attentional Deep Learning Model,” Mathematics, vol. 10, no. 24, Dec. 2022, doi: 10.3390/math10244733.

[7] F. E. Usman-Hamza et al., “Intelligent Decision Forest Models for Customer Churn Prediction,” Applied Sciences (Switzerland), vol. 12, no. 16, Aug. 2022, doi: 10.3390/app12168270.

[8] I. V. Pustokhina, D. A. Pustokhin, P. T. Nguyen, M. Elhoseny, and K. Shankar, “Multi-objective rain optimisation algorithm with WELM model for customer churn prediction in telecommunication sector,” Complex and Intelligent Systems, vol. 9, no. 4, pp. 3473–3485, Aug. 2023, doi: 10.1007/s40747-021-00353- 6.

[9] A. Amin et al., “Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study,” IEEE Access, vol. 4, pp. 7940–7957, 2016, doi:10.1109/ACCESS.2016.2619719.

[10] M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen, “Social Network Analytics for Churn Prediction in Telco: Model Building, Evaluation and Network Architecture,” Jan. 2020, doi:10.1016/j.eswa.2017.05.02

[11] C. P. Wei and I. T. Chiu, “Turning telecommunications call details to churn prediction: A data mining approach,” Expert Syst. Appl., vol. 23, no. 2, pp. 103–112, Aug. 2002

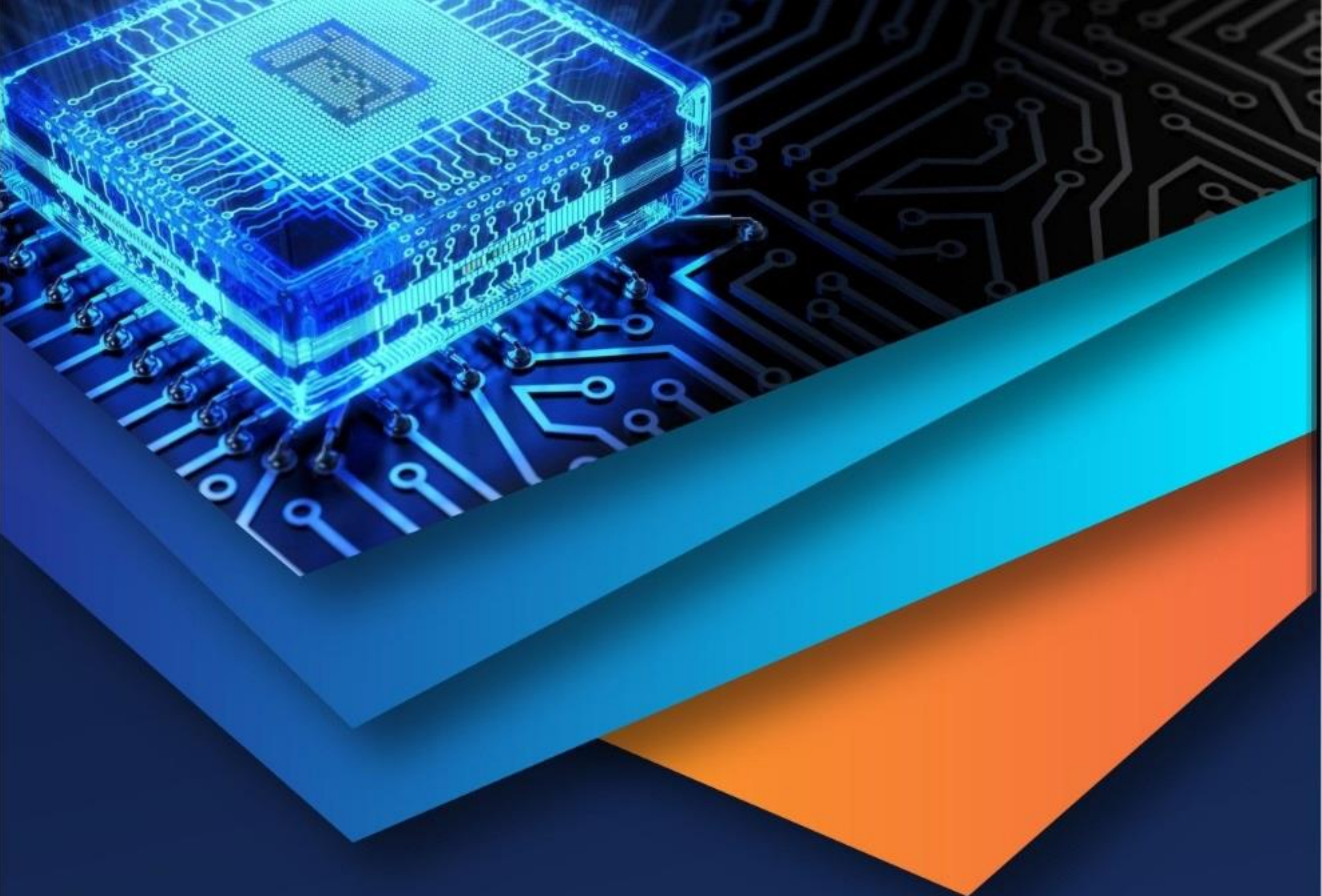
[12] S. Saleh and S. Saha, “Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university,” SN Appl Sci, vol. 5, no. 7, Jul. 2023, doi: 10.1007/s42452-023-05389-6.

[13] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, “Improved churn prediction in telecommunication industry using data mining techniques,” Applied Soft Computing, vol. 24, pp. 994–1012, 2014, 2014,doi:10.1016/j.asoc.2014.08.041

[14] N. Alboukaey, A. Joukhadar, and N. Ghneim, “Dynamic behavior-based churn prediction in mobile telecom,” Expert Syst. Appl., vol. 162, Dec. 2020, Art. no. 113779

[15] A. A. K. Ahmad, A. Jafar, and K. Aljoumaa, “Customer churn prediction in telecom using machine learning in big data platform,” J Big Data, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0191-6.

[16]



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY