An Improved Convolutional Neural Network for Churn Analysis

Priya Gopal, Dr. Nazri Bin MohdNawi
Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia

***Abstract*—The significance of customer churn analysis has escalated due to the increasing availability of relevant data and intensifying competition. Researchers and practitioners are focused on enhancing prediction accuracy in modeling approaches, with deep neural networks emerging as appealing due to their robust performance across domains. However, the computational demands surge due to the challenges posed by dimensionality and inherent characteristics of the data. To address these issues, this research proposes a novel hybrid model that strategically integrates Convolutional Neural Networks (CNN) and a modified Variational Autoencoder (VAE). By carefully adjusting the parameters of the VAE to capture the central tendency and range of variation, the study aims to enhance the effectiveness of classifying high-dimensional churn data. The proposed framework's efficacy is evaluated using six benchmark datasets from various domains, with performance metrics encompassing accuracy, f1-score, precision, recall, and response time. Experimental results underscore the prowess of the hybrid technique in effectively handling high-dimensional and imbalanced time series data, thus offering a robust pathway for enhanced churn analysis.**

Keywords—Customer churn analysis; deep learning; variational autoencoder; convolutional neural networks; dimensionality reduction

I. Introduction

In today's rapidly evolving business landscape, driven by the surge of online technological advancements, companies are compelled to navigate a competitive arena characterized by the influx of new business models and market entrants [1]. This has intensified the significance of customer churn analysis, as businesses seek to attract new customers and retain their existing clientele [2]. Retaining customers has been proven to yield higher returns on investment, as the costs associated with retaining an existing customer are considerably lower than acquiring a new one [3]. Amidst this context, the retention strategy gains paramount importance, requiring companies to mitigate the risk of customer churn – the phenomenon where customers switch providers swiftly [4-5].

To address this challenge, the utilization of machine learning has emerged as a potent tool, leveraging historical data to predict potential churn events and enable informed decision-making [6, 7]. However, there are hurdles to overcome in this endeavor. Issues such as inaccurate customer information, intricate datasets with numerous variables, imbalanced class distributions, and a lack of industry expertise create formidable hurdles [8]. Despite the strides made by advanced techniques like Convolutional Neural Networks (CNNs), which uncover hidden relationships within data, accurately predicting real-world churn scenarios remains intricate [9-11]. In light of these challenges, this paper introduces a hybrid model named the Space Vector Variational Autoencoder (SV-VAE), a fusion of CNN, and an optimized Variational Autoencoder (VAE) [12, 13].

By addressing these challenges, this study contributes to the enhancement of churn prediction in the dynamic landscape of modern business. It brings together cutting-edge technologies in a concerted effort to improve retention rates and elevate the strategic decision-making process for businesses across diverse industries.

The core objective of this paper is to enhance both the accuracy of predictions and the efficiency of model learning. This enhancement is achieved through the integration of a Convolutional Neural Network (CNN) with a modified Variational Autoencoder (SV-VAE). By combining these techniques, we aim to achieve superior performance in terms of predictive precision and reduced model training time.

To validate the effectiveness of the proposed hybrid model, a comprehensive evaluation is conducted. This evaluation encompasses various critical performance metrics, including precision, recall, accuracy, and learning time. To establish a robust baseline for comparison, the proposed SV-VAE hybrid model is benchmarked against other popular autoencoder architectures such as Vanilla, Stacked, Sparse, Denoising, and Variational Autoencoders. These comparisons are conducted across diverse industry-standard benchmark datasets, which provide a real-world context for assessing model performance.

The validation process primarily centers around the scrutiny of the proposed model's predictive capabilities The study meticulously assesses the accuracy of predictions, the ability to accurately classify positive instances (precision), and the model's effectiveness in capturing actual positive instances (recall). This thorough evaluation ensures that the proposed hybrid SV-VAE model's performance improvements are statistically significant and practically relevant in the context of churn analysis and prediction tasks. The following section contains a comprehensive review of the existing literature in the field of churn prediction, machine learning techniques, and autoencoder architectures relevant to this study.

II. Literature Review

Many methods have been explored in the quest to predict churn in service industries, often rooted in machine learning and data mining techniques. A significant portion of the prior research has been concentrated on individual data mining techniques or has involved comparative analyses of different methodologies for predicting attrition.

In a study conducted by Brandusoiu et al. [11], the focus was on predicting prepaid customer turnover rates using a contemporary data mining approach. The study utilized a dataset comprising more than 3000 call details, encompassing 21 attributes, and a predictive churn variable categorized with Yes/No labels. These attributes encompassed details about the volume of voice and video usage for each subscriber, alongside the count of inbound and outbound texts. The researcher employed the principal component analysis algorithm to streamline the data's complexity for dimensionality reduction. Three machine learning algorithms, namely Support Vector Machine (SVM), Naive Bayes (NB), and Neural Networks (NN), were employed to forecast churn rates. Model reliability was assessed using the Area under the Curve metric, and the results highlighted the superior performance of SVM over the other two algorithms. Notably, the dataset used in this study didn't contain any missing values. However, when dealing with time-series features, the model's ability to leverage information over time might be limited, necessitating various sampling techniques to incorporate temporal information effectively [12-13].

Artificial neural network approaches designed for sequential data have gained popularity, and this trend is evident in their increased adoption for churn modeling, as evident from the overview provided in Table I.

TABLE I. CHURN MODELS ANALYZED BY DIFFERENT AUTHORS

Paper	NN technique	Industry Data	Accuracy
Nasebah et al. (2019) [17]	CNN, Modified – CNN	Telecom	Accuracy precision, recall, F-measure, ROC & AUC
A. S. Kumara and D. Chandrakala (2016) [18]	LSTM, RFM + LSTM	Telecom	Mean evaluation metric
Domingos et al. (2021) [19]	MLP, DNN	Banking Sector	Accuracy using RMSProp, SGD ,Adam
Ahmed et al. (2019) [20]	CNN Classifiers, custom CNN	Telecom	Prediction Accuracy, ROC
Zhou et al. (2019) [21]	DL-CNN, One-dimensional CNN, XGBoost	Online New Media Platform	Precision Recall
Umayaparvathi and Iyakutti et al. (2017) [22]	CNN, mall FNN, Large FNN	Telecom	Accuracy
Wangperawong et al. (2016)	Deep CNN, Autoencoder	Time-series data	AUC
Kristensen et al (2019)	LSTM, Aggregated LSTM, LSTM Hidden State	freemium games	ROC, AUC & Accuracy
Prosvetov and Artem. (2018) [25]	CNN-based autoencoder, LSTM-based autoencoder	Telecom	roc-auc metric

Martins [14] conducted a study revealing that the accuracy of Long Short-Term Memory (LSTM) models equipped with time series attributes is comparable to an approach that integrates this pertinent data using the mean and a random forest technique. This research outcome contributes to synthesizing insights from various studies.

Numerous research endeavors have underscored the efficacy of combining Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) models to enhance performance across diverse tasks. For example, in a study by Tan et al., [15], a fusion of CNN and LSTM was employed to forecast user-intended actions, demonstrating that this integration surpasses the individual performance of the two models and traditional machine learning methods. Furthermore, another investigation employed encoded sequential data, such as images and videos, wherein the utilization of CNN yielded superior outcomes compared to gradient boosting and random forest techniques [16].

These discoveries accentuate the potential advantages of harnessing the strengths of distinct models, leading to heightened performance across a spectrum of applications.

In summary, previous studies indicate that the exploration of time-varying data to enhance the effectiveness of churn algorithms is still in its nascent stages. While CNN and LSTM models exhibit improved performance, they encounter challenges when confronted with high-dimensional data. Moreover, the diverse characteristics inherent to various industry sectors make it challenging to definitively determine the performance boost resulting from the inclusion of such varied information. Introducing these diverse features into the training phase can inadvertently amplify model complexity, potentially leading to overfitting against the training data. A potential solution to this lies in the preprocessing stage, where a dimensionality reduction step is undertaken. This step strives to curtail the number of features while retaining as much meaningful information as possible within the dataset [25]. Autoencoders prove adept at handling high-dimensional data, a domain where CNNs might face limitations. Notably, the Variational Autoencoder (VAE) is of special significance, given its ability to generate more probabilistic latent outputs [18]. The VAE emerges as a robust choice, particularly well-suited for churn analysis due to its capability to generate novel data instances and its compatibility with the time series nature of churn data. Additionally, dimensionality reduction plays a vital role in effectively mitigating noise from the data. This process facilitates the discovery of latent variables that arise from intricate relationships among different variables in the dataset. This approach provides a more comprehensive understanding that goes beyond analyzing individual variables in isolation.In conclusion, this study makes a valuable contribution to the services industry by meticulously assessing the efficacy of deep learning classification techniques and exploring alternative strategies to effectively manage high-dimensional data challenges.

III. RESEARCH METHODOLOGY

In scenarios involving time-varying features, various aggregation methods [23-24] have been explored alongside machine-learning classification techniques [26–27]. However, these methods often fall short due to the requirement of having one observation per client in most classification techniques. This limitation becomes problematic when tracking the behavior of the same customer over time with time-varying features. Consequently, conventional classification methods struggle to effectively utilize this type of information.

In order to address our problem effectively, we have devised a structured approach that leverages a modified Variational Autoencoder (VAE). This method aims to uncover latent space attributes, overcome challenges in traditional autoencoders, and generate new features from complex datasets.

Step 1: Variational Autoencoder (VAE) Setup

We begin by setting up a Variational Autoencoder (VAE), a powerful tool known for its ability to uncover latent space attributes in data. The VAE comprises two essential components: an encoder and a decoder.

Step 2: Encoder and Decoder Functions

The encoder processes input data samples and maps them to latent variables. This encoder is instrumental in generating meaningful latent features. On the other hand, the decoder strives to replicate the input data using the learned latent variable distribution.

Step 3: Leveraging Latent Space

Latent variables are relatively low-dimensional representations of the input data, which contrasts with the high-dimensional input and reconstructed data. This approach is built on the idea that data is generated by the model P(x|z).

Step 4: SV-VAE Architecture

Our proposed methodology incorporates four major blocks within the Space Vector Variational Autoencoder (SV-VAE): Encoder, Latent Distribution, KL Divergence, and Decoder. The SV-VAE leverages posterior distribution for data sampling and applies an empirical rule to reduce noise and approximate data points.

Step 5: Training the Model

Training involves optimizing two key loss functions: The KL divergence loss, which regularizes the learned latent distribution against a prior distribution, and the reconstruction loss, which ensures fidelity between decoded samples and original inputs.

Step 6: Deep Neural Network

Compressed features obtained from the SV-VAE are fed into a deep neural network with layers like pooling, dropout, ReLU, and a sigmoid layer. The output of this CNN flows into the decoder for data reconstruction.

Step 7: Evaluation

Model evaluation is performed through the assessment of SV-VAE loss, including the KL divergence loss function. Model predictions are evaluated for accuracy, F1-score, and precision. Hyperparameter tuning is carried out to enhance model accuracy.

Step 8: Dual Loss Optimization

Our SV-VAE model optimizes two crucial loss functions: The reconstruction loss, ensuring alignment with original dataset images, and the KL-divergence loss, quantifying the divergence from a standard normal distribution. This dual loss optimization ensures the model's effectiveness in capturing latent features.

The proposed model incorporates a modified Variational Autoencoder to uncover latent space attributes. VAE's ability to generate data across the entire space addresses the challenge of non-regularized latent space in traditional autoencoders. Within the VAE framework, an encoder module transforms the input sample x into a latent space representation x'. Variational autoencoders are particularly well-suited for generating new features from complex datasets [28].

The VAE consists of two core components: The encoder and the decoder. The encoder is a separate network that accepts samples from the data $\{x_i\}_{i=1}^{N}$ and attempts to map them to the latent variables z. The decoder, on the other hand, attempts to replicate the input $\{\hat{x}_i\}_{i=1}^{N}$ using the learned distribution z. Input x and reconstructed data samples x are in high dimensional space, however, latent variable z is relatively low dimensional. The foundation of the variational autoencoder rests on the notion that data is generated by the model P(x|z).

As illustrated in Fig. 1, the proposed methodology comprises four major blocks within the SV-VAE: Encoder, Latent Distribution, KL Divergence, and Decoder. The space vector variational autoencoder samples the data based on the posterior distribution. To remove noise and approximate data points, an empirical rule is applied. The encoder block plays a crucial role in generating latent features from the normal distribution data, emphasizing mean and standard deviation.

Training the model involves optimizing two loss functions: the KL divergence between the learned latent distribution and the prior distribution, which acts as a regularization term, and the reconstruction loss, which enforces fidelity between the decoded samples and the original inputs.

The compressed features from the SV-VAE are fed into a deep neural network that includes layers like pooling, dropout, ReLU, and a sigmoid layer. The output from the CNN flows to the decoder for data reconstruction. The evaluation of SV-VAE loss is accomplished through the KL divergence loss function. Similarly, the model's predictions are assessed for accuracy, f1-score, and precision. Hyperparameter tuning is conducted to enhance model accuracy.

The proposed SV-VAE model optimizes two key loss functions: reconstruction loss, which ensures that the decoder's output aligns with the original dataset images, and KL-divergence loss, which quantifies the divergence between the latent vector and a unit normal distribution. This divergence measurement ensures that the latent variables closely adhere to a standard normal distribution.

Martins [14] conducted a study revealing that the accuracy of Long Short-Term Memory (LSTM) models equipped with time series attributes is comparable to an approach that integrates this pertinent data using the mean and a random forest technique. This research outcome contributes to synthesizing insights from various studies.

Numerous research endeavors have underscored the efficacy of combining Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) models to enhance performance across diverse tasks. For example, in a study by Tan et al., [15], a fusion of CNN and LSTM was employed to forecast user-intended actions, demonstrating that this integration surpasses the individual performance of the two models and traditional machine learning methods. Furthermore, another investigation employed encoded sequential data, such as images and videos, wherein the utilization of CNN yielded superior outcomes compared to gradient boosting and random forest techniques [16].

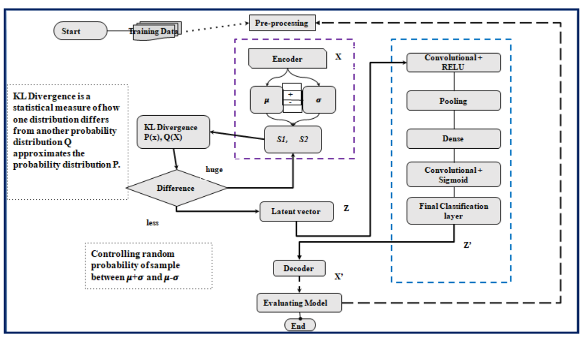

Fig. 1. Proposed SV-VAE model.

The sampling scenarios in VAE are to map the input to a distribution instead of mapping the input to a fixed vector.

$$x = sample\ (N(\boldsymbol{\mu},\boldsymbol{\sigma}\text{^}\mathbf{2}))$$

The modified approach uses mean and standard deviations to approximate the distribution of data,

$$x = (\mu+2\sigma) + (\ -2\sigma\ )\ sample(N(0,1))$$

IV. Exprimental Setup

A. Dataset

This study utilizes six distinct publicly available datasets from diverse domains, sourced from repositories like Kaggle and UCI. The datasets encompass a range of data types, including discrete, continuous, and categorical values. The dataset sizes vary, with a minimum of 954 records from the Tour & Travels domain to a maximum of 15,000 records from the Music streaming subscriptions domain. Additional insights regarding the dataset characteristics and the specifics of the training-test split are provided in Table II.

TABLE II. Detailed Churn Dataset

Domain	Number of records			No of attributes	Train Set	Test Set
	Churn	Non Churn	Total			
Bank	32000	68000	10000	14	80000	20000
Telecom – fixed line	2000	5000	70000	19	5600	1400
Employee churn	270	1200	1470	35	1176	294
Online subscription	15000	135000	150000	30	120000	30000
Tour & Travels	109	845	954	7	763	191
Telecom – mobile	869	2281	3150	13	2520	630

B. Hardware and Software

The study was conducted on an Ubuntu 20.04 LTS operating system, employing an i9 12th-generation processor coupled with 16GB RAM and a 1TB HDD. The implementation process was carried out using a Jupiter Notebook in Python v3.10.0. For the implementation, a suite of Python libraries was utilized, encompassing NumPy, Pandas, Seaborn, Sklearn, Keras, TensorFlow v2.0, and Matplotlib. These libraries played pivotal roles in both data pre-processing and modeling stages, contributing to the overall analysis.

C. Pre-Processing

Data pre-processing is a fundamental phase in the workflow of every machine learning engineer. This stage encompasses a range of essential steps aimed at refining the dataset for optimal analysis. Imputation of missing values, type conversion, duplicate removal, cleansing, normalization, and transformation are key procedures frequently applied in this phase. For addressing missing values, diverse strategies can be employed, such as statistical methods like mean, median, or even leveraging regression models to predict and fill in the absent values. Data cleansing, on the other hand, entails eliminating noisy data through techniques like binning, regression, and clustering. Once the crucial pre-processing steps are completed, a thorough analysis of the attributes follows, often leading to the creation of new features. The process of attribute selection involves assessing the correlation between variables and selecting the appropriate number of attributes that contribute most effectively to the analysis. To prepare the data for subsequent modeling, it is transformed into a structured format, typically in the form of two-dimensional arrays. These arrays are then divided into training and testing sets. The training data, which constitutes the input for model training, is carefully configured to enable accurate analysis and prediction.

D. Hyperparameter

The model is fine-tuned through the manipulation of hyperparameters, which play a crucial role in enhancing the algorithm's performance. These hyperparameters encompass attributes such as batch size, optimizer, number of epochs, learning rate, dropout rate, and random initialization. By carefully adjusting these parameters, the algorithm can be optimized to yield a more generalized and accurate model. Batch size, a vital hyperparameter in gradient descent, determines the number of training data instances utilized in each iteration. It governs the update of internal model parameters before proceeding to the subsequent iteration. The epoch parameter controls the iteration count for data feeding into the model. The learning rate hyperparameter adjusts the step size of weight adjustments during each epoch, critically influencing the optimization process. To guard against overfitting, a common challenge in model training, the dropout hyperparameter is introduced. This mechanism randomly omits a portion of neurons during training, preventing the model from becoming overly tailored to the training data. This practice enhances the model's capacity to generalize to unseen data.

In the context of this study, these hyperparameters are strategically manipulated to regulate the network's behavior during the training phase, ultimately contributing to the development of a more robust and efficient model.

V. Result and Discussion

The model's performance was assessed using essential metrics such as precision, recall, accuracy, and F1-score. Depending on the specific business context, the choice between prioritizing precision or recall was determined to gauge the effectiveness of the churn model. Each of these metrics is mathematically derived to provide a comprehensive understanding of the model's classification capabilities. These quantitative assessments serve as valuable tools for objectively evaluating the model's performance, catering to different business needs and objectives. Mathematically each of these measurements is derived by,

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{Total}$$

$$\text{F1} - \text{Score} = 2 * \left(\frac{Sensitivity * Specificity}{Sensitivity + Specificity}\right)$$

TP = True Positive FP = False Positive

FN = False Negative TN = True Negative

TABLE III. SV-VAE with Impact Dropout Over Accuracy

Dropout rate	0.1	0.25	0.5	0.75	0.9
Accuracy					
Telecom - fixed line	94	94.25	91.0	89.45	81.6
Telecom	96.2	96.15	90.4	85.6	78.0
Tour and travel	97.8	98.1	88.3	80.2	72.6
Banking	93.1	90.5	85.26	79.36	72.0
Music online subscription model	95.8	96.3	87.1	80.1	74.0
Employee retention	98.1	98.2	90.0	85.0	74.5

To investigate the behavior of the proposed model, a range of dropout values was experimented with (see Table III). It was observed that dropout values between 0.1 and 0.25 yielded optimal results, striking a balance between preventing overfitting and retaining useful information and Table IV shows the effect of different learning rates and average accuracy.

TABLE IV. Effect of Different Learning Rates and Avg. Accuracy

Learning rate	1	0.5	0.1	0.01	0.001	0.0001
Avg. Accuracy %	68.3	72.7	83.1	85.6	93.25	90.2

Table V displays the confusion matrix, offering a detailed breakdown of instances in which non-churn data is correctly classified as such (True Positives - TP) and churn data is accurately identified as churn (True Negatives - TN). In the context of a churn predictive model, the primary goal is the precise identification of churn users. This matrix provides a comprehensive assessment of the model's performance, shedding light on both accurate and erroneous classifications. Consequently, it informs the calculation of various evaluation metrics utilized in the analysis. In Fig. 2, a recall comparison between the proposed model and standard autoencoders is presented, highlighting the superior recall performance of the proposed model.

TABLE V. Model Evaluation with Confusion Matrix - Online Music Streaming Subscription Dataset

	Actual Churn	Actual Not Churn
Predicted Churn	**4300**	**300**
Predicted Not Churn	**200**	**145000**
Total Records	**150000**	
Total Not Churn	**148000**	
Total Churn	**4500**	
Precision	**0.9712**	
Recall	**0.9966**	
Accuracy	**0.9680**	

The evaluation of recall enhancement not only underscores the technical progress achieved with the SV-VAE model but also carries profound implications for churn analysis. The notable uptick in the average recall, approximately 4.38%, signifies a substantial boost in the model's capability to accurately detect instances of churn. Within the realm of churn analysis, recall serves as a pivotal metric, quantifying the model's proficiency in capturing genuine churn occurrences among the overall churn cases. This enhancement directly translates into a more potent identification of customers at risk of churning, thereby equipping businesses with the proactive means to intervene and retain these valuable customers.

Table VI presents a comparative analysis of churn prediction accuracy using different types of autoencoders across various domains. Notably, the SV-VAE model consistently stands out, demonstrating superior accuracy across multiple sectors. The remarkable improvement of 5.01% in average accuracy underscores the model's enhanced ability to make precise classifications, distinguishing between churn and non-churn instances with greater accuracy. In churn analysis, accuracy is a vital metric that quantifies the overall correctness of the model's predictions. The improved accuracy ensures that the decisions based on the model's predictions are more reliable, leading to optimized resource allocation for customer retention efforts and yielding better business outcomes. In both instances, these improvements in recall and accuracy substantiate the efficacy of the SV-VAE model in the realm of churn analysis. By accurately identifying potential churners and enhancing overall classification precision, the SV-VAE model enables businesses to devise more targeted and effective strategies to mitigate customer churn. This not only contributes to retaining valuable customers but also optimizes resource allocation and strategic decision-making, ultimately bolstering the competitive edge of businesses in the market.

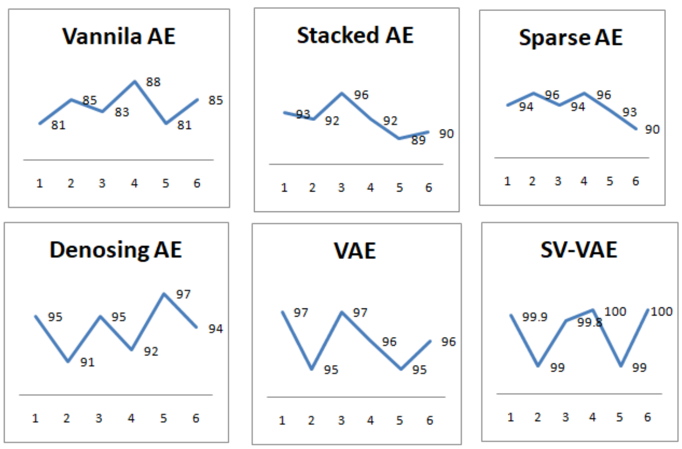

Fig. 2. Recall of the different types of autoencoders with proposed SV-VAE.

TABLE VI. CHURN PREDICTION ACCURACY OVER VARIOUS DOMAINS COMPARED AGAINST DIFFERENT TYPES OF AUTOENCODERS

Model/Dataset	Vanilla AE	Stacked AE	Sparse AE	Denoising AE	VAE	SV-VAE
Telecom Mobile	81	93	94	95	97	99.9
Bank	85	92	96	91	95	99
Music streaming	83	96	94	95	97	99.8
Employee	88	92	96	92	96	100
Telecom – Fixed line	81	89	93	97	95	99
Tour & Travels	85	90	90	94	96	100

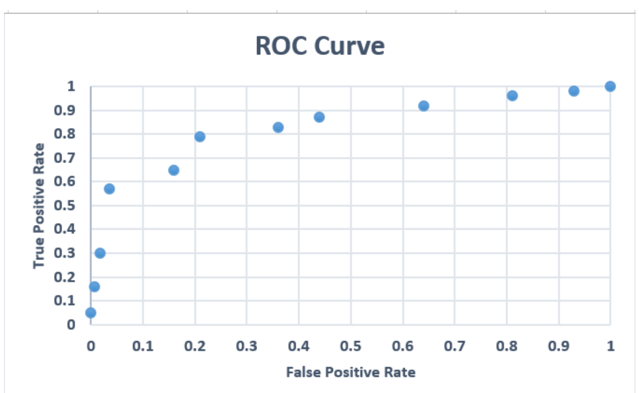

Fig. 3. ROC Curve of the proposed SV-VAE.

A ROC (Receiver Operating Characteristic) curve is a graphical tool that shows how the True Positive Rate (TPR) and False Positive Rate (FPR) change when we adjust the threshold for classifying data points as either positive or negative.

$$FPR = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$$

$$\text{TPR} = \frac{\text{True Positives}}{\text{False Positives} + \text{True Negatives}}$$

By adjusting this threshold, one can observe the variation in TPR (True Positive Rate) and FPR (False Positive Rate) values. Typically, as the threshold decreases, TPR increases, but FPR also rises. The ROC (Receiver Operating Characteristic) curve provides a visual representation of this trade-off and serves as a valuable tool for assessing a model's performance across various thresholds. The AUC (Area Under the Curve) is a quantitative metric that summarizes the model's overall performance over all conceivable thresholds. It quantifies the model's capacity to differentiate between positive and negative instances. In the case of the proposed model, the AUC value stands at 92.45. A greater AUC score denotes enhanced discriminatory power, with a value of 1 denoting a model that operates perfectly, and 0.5 indicating a model that merely makes random guesses. Fig. 3 graphically presents the ROC curve of the proposed model, offering a clear visual representation of its discriminatory power.

VI. CONCLUSION AND FUTURE WORK

The proposed hybrid model, known as the Space Vector Variational Autoencoder with Convolutional Neural Networks (SV-VAE with CNN), represents a powerful solution for churn prediction tailored to the unique characteristics of churn data. The proposed approach represents a significant departure from prevailing systems in several key aspects. While contemporary systems often rely on traditional machine learning techniques and struggle to effectively utilize time-varying features, this method harnesses the power of a modified Variational Autoencoder (VAE) to unlock latent data attributes. This approach offers several distinctive advantages. It excels in the effective handling of time-varying features. Unlike conventional systems that face limitations when tracking the behavior of the same customer over time with time-varying features, this approach excels in this regard. By leveraging a VAE, it can capture the dynamic nature of features and generate latent representations that encapsulate temporal patterns. Additionally, it effectively reduces data dimensionality through the VAE's latent space, enabling more efficient analysis and modeling. In comparison to standard autoencoders, this approach incorporates Space Vector Variational Autoencoder (SV-VAE) architecture, enabling better discrimination and noise reduction, contributing to more accurate predictions. Moreover, while some systems focus solely on reconstruction loss, this approach optimizes two critical loss functions: reconstruction loss and KL-divergence loss, ensuring the effective capture of latent features while adhering to a standard normal distribution.

This research also opens avenues for generalization to more intricate scenarios and challenges.

Multi-Modal Data Integration: The approach, rooted in the VAE framework, can readily adapt to scenarios involving multi-modal data sources. By extending the encoder and decoder components, it can incorporate various data types and establish a more comprehensive understanding of complex cases.

Temporal Sequence Modeling: While addressing time-varying features, there is potential to explore more advanced temporal sequence modeling techniques, such as incorporating recurrent neural networks (RNNs) or attention mechanisms to capture intricate temporal dependencies.

Transfer Learning and Scalability: As the foundation of this approach lies in feature extraction and latent space representation, it is well-suited for transfer learning, allowing for the application of knowledge gained from one domain to another. Additionally, this methodology can be scaled to accommodate larger datasets and more extensive feature sets by leveraging distributed computing and parallel processing, extending its applicability to handle big data scenarios.

In conclusion, the proposed approach not only distinguishes itself from existing systems but also paves the way for broader applications in complex cases. These differences and potential generalization pathways are discussed here to provide a more comprehensive view of the research's contributions and future possibilities.

References

[1] Gerpott TJ, Rams W, Schindler A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommun Policy. 2001;25:249–69.

[2] Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: a data mining approach. Expert Syst Appl. 2002;23(2):103–12.

[3] Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. 2013. p. 131–6.

[4] Ascarza E, Iyengar R, Schleicher M. The perils of proactive churn prevention using plan recommendations: evidence from a field experiment. J Market Res. 2016;53(1):46–60.

[5] Bott. Predicting customer churn in telecom industry using multilayer perceptron neural networks: modeling and analysis. Igarss. 2014;11(1)

[6] Umayaparvathi V, Iyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods, and metrics. Int Res J Eng Technol. 2016;3(4):1065–70.

[7] Yu W, Jutla DN, Sivakumar SC. A churn-strategy alignment model for managers in mobile telecom. In: Communication networks and services research conference, vol. 3. 2005. p. 48–53.

[8] Mena, C.G., Caigny, A.D., Coussement, K., Bock, K.W., & Lessmann, S. (2019). Churn Prediction with Sequential Data and Deep Neural Networks. A Comparative Analysis. ArXiv, abs/1909.11114.

[9] De Caigny, A., K. Coussement, K. W. D. Bock, and S. Lessmann (2019): "Incorporating textual information in customer churn prediction models based on a convolutional neural network," International Journal of Forecasting.

[10] Allam, Swetha. (2019). Churn Prediction using Attention Based Autoencoder Network. International Journal of Advanced Trends in Computer Science and Engineering. 8. 725-730.

[11] Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

[12] Wei, C.-P. and I.-T. Chiu (2002): "Turning telecommunications call details to churn prediction: a data mining approach," Expert Systems with Applications, 23, 103 – 112

[13] Song, G., D. Yang, L. Wu, T. Wang, and S. Tang (2006): "A Mixed Process Neural Network and its Application to Churn Prediction in Mobile Communications," in Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), 798–802.

[14] Martins, H. (2017): "Predicting user churn on streaming services using recurrent neural networks,"

[15] Tan, F., Z. Wei, J. He, X. Wu, B. Peng, H. Liu, and Z. Yan (2018): "A, Blended Deep Learning Approach for Predicting User Intended Actions," 2018 IEEE International Conference on Data Mining (ICDM), 487–496.

[16] Zaratiegui, J., A. Montoro, and F. Castanedo (2015): "Performing Highly Accurate Predictions Through Convolutional Networks for Actual Telecommunication Challenges," CoRR, abs/1511.04906.

[17] Nasebah Almufadi, Ali Mustafa Qamar, Rehan Ullah Khan, Mohamed Tahar Ben Othman, Deep Learning-based Churn Prediction of Telecom Subscribers, International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 2743-2748

[18] Joolfoo, Muhammad. (2020). Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform.

[19] Domingos, Edvaldo & Ojeme, Blessing & Daramola, Olawande. (2021). Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. Computation. 9.0.3390/computation9030034.

[20] Ammar A.Q. Ahmed, D. Maheswari, Churn prediction on huge telecom data using hybrid firefly based classification, Egyptian Informatics Journal, Volume 18, Issue 3,2017,Pages 215-220,ISSN 1110-8665,https://doi.org/10.1016/j.eij.2017.02.002.

[21] Wang, Li and Chen, Chaochao and Zhou, Jun and Li, Xiaolong(2018) Time-sensitive Customer Churn Prediction based on PU Learning, arXiv, 10.48550/ARXIV.1802.09788

[22] Umayaparvathi, V. & Iyakutti, K.. (2012). Applications of Data Mining Techniques in Telecom Churn Prediction. International Journal of Computer Applications. 42. 5-9. 10.5120/5814-8122.

[23] Wangperawong, Artit & Brun, Cyrille & Laudy, Olav & Pavasuthipaisit, Rujikorn. (2016). Churn analysis using deep convolutional neural networks and autoencoders.

[24] Kristensen, Jeppe & Burelli, Paolo. (2019). Combining Sequential and Aggregated Data for Churn Prediction in Casual Freemium Games. 1-8. 10.1109/CIG.2019.8848106.

[25] Prosvetov, Artem. (2018). The comparison of autoencoder architectures in improving of prediction models. Journal of Physics: Conference Series. 1117. 012006. 10.1088/1742-6596/1117/1/012006.

[26] C. Wei and I. Chiu, Turning telecommunication call details to churn prediction: a data mining Approach expert System with applications, 2002.

[27] Y. Liu, Z. Xu, J. Yang, L. Wang, C. Song and K. Chen, "A Novel Meta-Heuristic-Based Sequential Forward Feature Selection Approach for Anomaly Detection Systems," 2016 International Conference on Network and Information Systems for Computers (ICNISC), 2016, pp. 218-227, doi: 10.1109/ICNISC.2016.056.

[28] Singh A, Ogunfunmi T. An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications. Entropy (Basel). 2021 Dec 28;24(1):55. doi: 10.3390/e24010055. PMID: 35052081; PMCID: PMC8774760.