IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 August 2023, accepted 17 November 2023, date of publication 23 November 2023, date of current version 11 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3336424

RESEARCH ARTICLE

New Hybrid Deep Learning Models to Predict Cost From Healthcare Providers in Smart Hospitals

MUHAMMAD HAMZA RAFIQUE BHATTI[1], NADEEM JAVAID[1], (Senior Member, IEEE), BABAR MANSOOR[2], NABIL ALRAJEH[3], MUHAMMAD ASLAM[4], AND MUHAMMAD ASAD[5], (Member, IEEE)

[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan
[2]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad 44000, Pakistan
[3]Department of Biomedical Technology, College of Applied Medical Sciences, King Saud University, Riyadh 11633, Saudi Arabia
[4]Department of Computer Science, Aberystwyth University, SY23 3FL Aberystwyth, U.K.
[5]Department of Creative Informatics, School of Information Science and Technology, The University of Tokyo, Tokyo 113-8654, Japan

Corresponding author: Nadeem Javaid (nadeemjavaidqau@gmail.com)

This work was supported by King Saud University, Riyadh, Saudi Arabia, through the Researchers Supporting Project under Grant RSP2023R648.

ABSTRACT Accurate cost prediction of healthcare resources is challenging as diverse factors affect the overall prediction. The cost of healthcare providers is increasing exponentially as different healthcare providers charge differently for the same service due to various factors, majorly the sky rocketing inflation and increased population. It increases the importance of predicting healthcare costs to avoid unpleasant surprises. This study aims to provide the expected cost of healthcare providers that helps the patients in resource allocation and strengthens decision-making according to their resources. This paper proposes three hybrid Deep Learning (DL) models, Visual Geometry Group and Stacked Autoencoder (VGG-SAE), Visual Geometry Group and Deep Neural Network (VGG-DNN), and Stacked Autoencoder and Deep Neural Network (SAE-DNN), which optimize learning the hidden patterns from the given data more efficiently than individual models. The three hybrid DL models estimate the cost of healthcare providers effectively. The preprocessing is performed using the mode imputation for handling the missing values, Z-score for removing the outliers and standard scaler for standardizing the data. To train the hybrid models on optimum parameters, the Random Search technique is used that provides the best hyper-parameters of each hybrid model. The interpretation of the hybrid models' output is achieved using the SHapley Additive ExPlanations (SHAP) technique. The performances of VGG-SAE, VGG-DNN, and SAE-DNN are compared with the baseline DL models such as SAE, DNN, and VGG. To assess the robustness of the proposed approach, the hybrid models are trained on two different datasets of healthcare such as Healthcare Providers and Hospital Inpatient Cost Transparency. With the hyper-parameter tuning of the Healthcare Providers Dataset, VGG-SAE achieved MSE of 0.01, RMSE of 0.13, MAE of 0.02, and R-squared of 0.98. VGG-DNN achieved MSE of 0.01, RMSE of 0.12, MAE of 0.02, and R-squared of 0.99. SAE-DNN achieved MSE of 0.01, RMSE of 0.11, MAE of 0.02, and R-squared of 0.99. With the hyper-parameter tuning of the Hospital Inpatient Cost Transparency Dataset, VGG-SAE achieved MSE of 0.007, RMSE of 0.08, MAE of 0.03, R-squared of 0.99, and execution time of 1680 seconds. VGG-DNN achieved MSE of 0.0006, RMSE of 0.08, MAE of 0.03, R-squared of 0.99, and execution time of 645 seconds. SAE-DNN achieved MSE of 0.003, RMSE of 0.06, MAE of 0.02, R-squared of 0.99, and execution time of 850 seconds. Our proposed hybrid combinations outperformed other deep models and Machine Learning (ML) techniques such as SAE, DNN, VGG, SVR and GBR, which ensures high efficiency of the proposed models in terms of healthcare providers cost.

INDEX TERMS Cost prediction, deep learning, healthcare, hybrid model, preprocessing, smart hospitals.

The associate editor coordinating the review of this manuscript and approving it for publication was Tai-Hoon Kim.

I. INTRODUCTION

Healthcare services are one of the main components of welfare systems. Due to their vast applications, they prove beneficial for both public and private sectors. Thus, they are one of the major consumers of the state accounts. Healthcare providers deliver their services in multiple fields to facilitate the patients like providing interventions and procedures to diagnose the patient's diseases. Health services are a necessity and are defined as the price of the utilization of goods and services related to healthcare. This healthcare consumption is related to different activities like outpatient, inpatient, diagnostic tests, surgeries, etc., [1]. In the United States, the expenditures in the health sector were almost 18% of the annual gross domestic product in 2017 [2]. With the passage of time, the requirement of healthcare services exponentially increases, which results in increase in the cost of healthcare. It depends on different associated factors, which further impact the cost utilization of healthcare services like comorbidity, patient parameters, and medical needs [3]. Healthcare utilization specifically depends on the severity and health condition of the patients as the cost of healthcare varies on the basis of patients' critical health aspects. Healthcare providers provide services in different domains such as health finance, providers and health tech. Healthcare providers are proven to have expertise in different fields of health. Besides, they exhibit heterogeneous characteristics. Healthcare utilization is a non-linear relationship that depends on the severity of the health conditions of patients. Cost prediction in the healthcare sector is a challenging task as various healthcare related factors such as the location of the providers, surgeries, pharmacy and labs affect the outcomes of the model. The cost prediction facilitates the patients in estimating the expected healthcare utilization that helps in the decision-making of the utilization of health costs [4]. It also helps the patients to predict the billing expenditure to avoid the unexpected cost of healthcare and efficiently manage the resources such as insurance policies and financial outcomes. The cost of healthcare services is very grinding to evaluate as each patient may have a different level of healthcare utilization. Also, the utilization intensity of a specific health service is different from others. Moreover, the same healthcare services have different prices in different healthcare hospitals as they are influenced by different factors such as costly equipment, location of hospital and the economic factors [1], [5].

The utilization of healthcare services is high in the old age patients. The healthcare providers are working intensely day and night. The number of patients are exponentially increasing, which results in the generation of healthcare data in large volumes, which acts as the main factor for cost prediction of healthcare providers. By utilizing the healthcare providers data, the useful information is extracted by performing multiple preprocessing techniques to find the patterns in the data and assist in decision-making related to the health services' utilization [1], [6]. With the emergence of Artificial Intelligence (AI), healthcare data is easily extracted, and different statistical operations are performed to evaluate the predictions in an effective way. It ensures continuity in the outcomes by building the decision-making criteria for the health services. AI is utilized in the field of healthcare to predict the cost expectations related to the healthcare services [6].

With the availability of a large volume of healthcare providers data, Machine Learning (ML) exhibits the form of AI that easily maps the relevant data and its variables to predict the patient's cost more precisely and accurately than traditional methods. ML finds the relationship between the inputs and outputs of the relevant data, trains the model on specific data and finally makes a decision on the basis of available data [7], [8], [9]. It improves the accuracy and precision by learning the hidden patterns in the training data [10], [11], [12], [13], [14], [15]. AI trains the DL models based on advanced mathematical formulation and identifies the patterns in the available data [1], [16], [17].

DL is the subset of AI that learns the hidden patterns from the training data more accurately than the traditional ML methods. Due to the massive amount of available data, DL models perform more perfectly as compared to ML algorithms. The learning patterns of healthcare providers are depicted by applying different DL models, finding the relationship between the non-linear data and predicting the cost of healthcare providers of different diagnoses relative to the expenditures and severity of patients [1], [18]. The major challenge in DL models that makes prediction less accurate and inefficient is the non-processing of data as it includes noisy values and missing data. It affects the performance of predictive models and increases the error in regression. Also, it considers the outliers present in the datasets that increase the factor of over-fitting in the prediction. It also estimates the undesirable results, which affect the overall performance of the model. The output of DL models exhibits the black box approach in which internal working is hidden. In order to interpret the internal working of DL models, SHAPley Additive ExPlanations (SHAP) technique is applied to explain the importance of each feature [3], [10], [19], [20], [21], [22].

A. CONTRIBUTIONS

The main contributions of this paper are as follows.

- Preprocessing techniques such as Standard Scaler, Z-Score and mode imputation are used to refine the data.
- Random search is used to find the optimal hyper-parameters of each DL model.
- DL models are used to mitigate the curse of dimensionality.
- Visual Geometry Group and Stacked Autoencoder (VGG-SAE), Visual Geometry Group and Deep Neural Network (VGG-DNN), and Stacked Autoencoder and

Deep Neural Network (SAE-DNN) hybrid models are proposed for cost prediction of healthcare providers.
- Mean Squared Error (MSE), Mean Average Error (MAE), Root Mean Squared Error (RMSE), and R-squared performance metrics are used for performance evaluation.
- SHAPley technique is used to explain the outcome of each DL model.

The remaining paper exhibits the following structure. In Section II, the related work is presented, and the Problem statement is elaborated in Section III. The system model is elaborated in Section IV. We discussed the simulation results in Section V, and finally, the conclusion is stated in Section VI.

TABLE 1. List of abbreviations.

Abbreviation	Definition
AI	Artificial Intelligence
AE	Autoencoder
ANN	Artificial Neural Network
BGR	Boosting Gradient Regressor
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
FC	Fully Connected Network
GBT	Gradient Boosting Trees
K-NN	K-Nearest Neighbor
LR	Linear Regression
LSTM	Long Short Term Memory
ML	Machine Learning
MAE	Mean Absolute Error
MSE	Mean Squared Error
MLP	Multi-Layer Perceptron
NB	Naive Bayes
RMSE	Root Mean Squared Error
SAE	Stacked Autoencoder
SVM	Support Vector Machine
SVR	Support Vector Regessor
VGG	Visual Geometry Group

II. RELATED WORK

In this section, the previous studies related to healthcare are discussed. Most related healthcare papers are grouped and classified into separate categories. This section is divided into two categories: cost prediction and disease prediction.

A. COST PREDICTION

Sanz et al. proposed a framework in which Artificial Neural Network (ANN) and Deep Neural Network (DNN) models were used to accurately estimate the billing prediction of healthcare. The healthcare sector generates a large amount of data, which is significantly big in volume and can be used for predicting the healthcare cost. The main challenge in healthcare is financial management as same services have different costs across the same city. However, healthcare prediction is important to avoid unpleasant surprises. The data was obtained from the public dataset and multiple algorithms were applied and compared with the Decision Tree (DT) to evaluate the performance of the model in [1]. In [22], Teo et al. addressed the issue of readmission rates in hospitals and proposed a solution to reduce the readmission rate. Readmission occurs due to the negligence of the healthcare providers and due to the increase in readmission rate. The reduction program was introduced that reduced the payments made in hospitals. It increased the readmission rate within 30 days. To tackle the issue of readmission rate, predictive solutions were proposed using AI to reduce the rate that also decreases the healthcare cost. Convolution Neural Network (CNN) based models were used to develop predictive models that predicted patients with high needs or high utilizers, which can further suggest an intervention to deal with the reduction in readmission rates and provide quality healthcare services.

Madison et al. [2] highlighted the high utilizers problem in healthcare sector. Most of the healthcare services are used and costs are incurred by the high utilizers. High utilizers were not properly identified and interventions were not followed that resulted in rise of healthcare expenditure. According to the survey, around 55 percent of healthcare costs were consumed by 5 percent of patients. Identifying these patients decreases the healthcare expenditure. The authors used unsupervised clustering approach to identify the patients who undergo Total Hip Arthroplasty (THA) or Total Knee Arthroplasty (TKA) to identify the patterns of utilization. Accurate prediction is a challenging aspect specifically in regression problems when the output is continuous. Norat et al. proposed a solution in which Medicare payments to be made to physical therapists were elaborated. The number of Americans with a demand for physical therapists has exponentially increased over the past 65 years. So, the prediction of accurate payments to therapists was a big challenge as the local factors also influenced the final outcomes. The genetic algorithm was proposed with the self-adaptation concept in which parameters of genetic algorithm were tuned automatically without the involvement of the human factor that increases the overall accuracy. It also tends to overcome the over-fitting problem by incorporating the self-adaptation in [23]. Berger et al. highlighted the high utilizers disease, Critical Limb Ischemia (CLI), that increases healthcare utilization and the chance of re-admission. Using the Bayesian ML platform, all-cause hospitalization was estimated and the overall annual healthcare cost was predicted. The authors used the patients' historical data and predicted the high predictors of all-cause hospitalization. The proposed model helped the patients to identify the disease related to the CLI and to predict the cost related to the CLI in [4].

Luo et al. [3] elaborated on the comorbidity problem in asthma patients, which leads to an increase in healthcare cost utilization. Accurate prediction of healthcare costs is always a challenging issue in asthma patients as these patients were affected by different comorbidity that makes accurate prediction of cost a difficult process. The comorbidity factor also depends on the severity of the disease in asthma patients. However, the comorbidity portfolio design leads to accurate prediction of cost. Different ML algorithms were trained to predict the overall cost by considering the data from 2012 to 2014 acquired from a large city in China. Different risks related to asthma patients were highlighted and discussed like respiratory disease and circulatory disease that can highly impact the prediction of cost in asthma patients.

Predicting the length of stay is always a critical issue as multiple factors are associated that influence the prediction. In [10], the authors highlighted the prediction problem in shoulder arthroplasty. The prediction of the expected healthcare cost utilization and length of stay were based on severity of the disease is quite challenging. The authors proposed ANN to accurately predict the length of stay, inpatient cost and discharge disposition by considering the data from 2003 to 2014 related to shoulder arthroplasty. The preoperative prediction helps the patients to estimate the overall expenses, cost related to the disease and length of stay based on the intensity of disease.

In [24], Morid et al. highlighted the effective cost prediction problem. The authors proposed a CNN model to automatically learn the hidden features from the multi-variate time series data using three years of medical and pharmacy claims data related to a patient's health status, visit, and cost features acquired from a healthcare insurer. The hyper-parameter of the proposed CNN model was also tuned to best fit the hidden patterns. The CNN architecture estimated the cost of individual patients by mapping the dependent and independent features. CNN comprises of three convolution layers and a pooling layer with Leaky Rectifed Linear Activation (LReLU) function having the customized value of kernel on each layer of the proposed model.

Estimating patient expenses is always a challenging task because different factors and severity of diseases affect the accurate prediction of patient expenses. ZENG et al. proposed the multi-view DL framework that used the heterogeneous patient historical claims data from January 2013 to December 2014 such as patient demographic features, medical codes, drug usage and facility utilization. It was trained on this data to predict individual expenditure precisely. The model comprised of different input sequences and inputs were passed to multiple DL models. The demographic features were passed to the feed-forward neural network, the utilization sequence was forwarded to the attention-based bi-directional RNN, and the medical code sequence was forwarded to the stacked RNN. Finally, the results were commuted to accurately predict the patient's individual expenditure [25].

In the healthcare sector [26], prediction of medicine spending is a critical issue. Different factors affect the prediction of medicine. Kaushik et al. proposed using the Variance-Based Generative Adversarial Network (V-GAN) model on patient data for prediction of pain medication. Different ML algorithms were used for prediction such as Multi-Layer Perceptron (MLP), CNN, and Long Short Term Memory (LSTM). However, the GAN was not used for medicine prediction. The GAN comprises of generator and discriminator. LSTM was used as a generator and CNN or MLP was used as a discriminator. In the proposed model, the V-GAN aimed to reduce the variance between actual and predicted values of the training data. V-GAN was compared with other GAN variants and ML models such as Linear Regression (LR) and Gradient Boosting Regressor (GBR). V-GAN outperformed LSTM and CNN in helping the patients to accurately predict the expected expenditure of medicine [26]. Making predictions of in-hospital resource consumption was quite challenging. The length of stay, inpatient and outpatient cost, and surgeries can affect the results of predictions. Yu et al. [27] predicted the cost of patient spending at the hospital from admission records. It contained heterogeneous data such as the patient features and diagnosis texts. The proposed model used the transformer to predict the patient cost by considering the heterogeneous data. Transformer was utilized to capture the representation of words, diagnosis and operations. Besides this, the diagnosis-operation mechanism was developed to find the relationship between diagnosis and operations. The authors also incorporated the hierarchical attention network that aimed to find the important word, diagnosis and operations for learning the representation of patient information [27].

In healthcare sector, some surgical procedures are very expensive and add to the patient's total bill such as TKA. Abbas et al. highlighted TKA as a resource extensive procedure. The cost mainly depended on two factors that were Duration Of Surgery (DOS) and length Of stay after the operation. So, the authors predicted DOS and length Of stay based on preoperative factors. The authors used multiple ML algorithms such as linear, tree-based and MLP. The models evaluated the accuracy of prediction using the data of national surgical, quality improvement and American College of Surgeons from 2014 to 2019. MLP performed the best among all ML methods. It predicted the DOS and length Of stay very accurately that helped the patients for efficient resource allocation [28].

B. DISEASE PREDICTION

In [29], Puri et al. overcame the issue of inefficient disease prediction using a Gaussian process-based model. It exhibited characteristics to train the model in the absence of insufficient data and predict the data accurately. Unavailability of data in the healthcare sector due to the number of subjects might be less or the data is obtained at a very low sampling frequency. The authors also provided a novel approach for the subset selection technique, which selected the time series data that matches the temporal equality with the times of interest. In [30], the diagnosis of heart disease was performed using AI in the healthcare sector. Early identification of heart disease plays an important role in the field of cardiology. Li et al. suggested an approach in which different ML algorithms were used to classify whether the patients are suffering from heart disease or not. Different algorithms such as Support Vector Machine (SVM), LR, ANN, K-Nearest Neighbor (K-NN), Naive Bayes (NB), DT, and different combinations of feature selection techniques were also used with these algorithms. The authors also proposed the feature selection technique, Fast Conditional Mutual Information (FCMIM) that performed better than traditional feature selection techniques. The proposed technique showed that with SVM accurate diagnosis of heart disease was ensured.

During COVID peak time in 2020, the physical interaction between doctors and patients exponentially reduced. Due to isolation, the interaction was performed via virtual sessions through the Internet. Due to the unavailability of physical doctors, patients tend to self diagnose their disease by searching on the Internet, reading multiple blogs that misguided and wrongly interpreted patients. To overcome this critical issue, Desai et al. developed an efficient solution that provided the best explanatory details regarding the patient's specific disease. It predicted whether the patient has a disease or not by training the model on health disease data. Multiple ML algorithms were used to predict heart disease that are SVM, K-NN, neural networks, LR, and Gradient Boosting Trees (GBT). The model proposed in [31] accurately predicted heart disease in patients.

Healthcare resource utilization in poor countries suffers due to unavailability of resources and infrastructure for healthcare. Lung cancer is one of the critical and the deadliest diseases. Health-related programs were conducted to treat lung cancer patients at a very early stage. The emergence of AI and Internet of Things (IoT) solved many complex issues. It helped healthcare providers in ensuring better meditation and performing an early prediction of disease. In poor regions where resources are insufficient to handle complex medical applications, Gu et al. suggested a solution in which medical industry 4.0 was utilized to overcome the issue of unavailability of resources. The proposed model was deployed on cloud platforms. The doctors easily predicted whether the patient was affected by lung cancer or not. The doctors also compared the predicted results with other similar cases for better medications and treatment of the patients. The model proposed in [32] helped the healthcare providers with decision-making regarding patient treatment. Emerging technologies also change the trends in healthcare industry by incorporating the latest technologies and frameworks to help healthcare providers improve health services.

Qureshi et al. utilized cloud infrastructure to revolutionize the medical industry with the latest technologies. The proposed solution consisted of sensors that aimed to sense the patient information, collect the information and pass it to the local datasets, and propagate the collected information to the cloud datasets via the cellular networks. ML techniques were also being deployed on the cloud platform. These techniques were trained using the data stored in the cloud dataset. A secure mobile-based solution was proposed in [33] in which information was collected from the patient and was propagated to the cloud. Further, the information was passed to the ML algorithms that were responsible to classify the cardiovascular diseases. Cheon et al. raised the issue of stroke in the Korean population. With the aging effect, the chance of stroke increases, which heavily increases the healthcare resource consumption. Authors proposed a model in which DNN was used to predict stroke outcomes in patients by incorporating the medical service and health behavior data. Before prediction with DNN, Principal Component Analysis (PCA) was incorporated to find the important features from the data. It improved the outcomes of the model as only needed information was fed to the predictive model [34].

Tripoliti et al. implemented multiple ML techniques such as K-NN, MLP, self-organizing maps, Classification And Regression Trees (CART), RF, SVM, neural networks, LR, DT, clustering and Fuzzy genetic to diagnose heart failure, and to predict the mortality and re-hospitalization. Heart failure is a critical condition that leads to two percent of total health costs in developed countries. Early prediction of heart failure improves the health of patients and minimizes the patient hospitalization cost [35]. Bhuyan et al. [36] proposed a generic model for identifying diseases, which is a challenging task. In healthcare systems, the most difficult and important task is to accurately identify the disease. Most of the healthcare providers resources were utilized for diagnosis of a disease in which the healthcare providers suggest the patients to undergo different methods, which help the doctors in making decisions regarding proper diagnosis of the patients. Different ML algorithms such as SVM, K-NN, RF, and LR were used to identify the disease. Among all these algorithms, ANN performed the best based on its accuracy and different evaluation parameters [36]. Prediction of time series data requires a large volume of data, and the unavailability of this time series data is a challenging aspect to train the model with insufficient data.

III. PROBLEM STATEMENT

Estimating healthcare services and utilization cost is a critical challenge as many factors affect the outcomes of patients. It makes it difficult to analyze the outcomes of healthcare services [2]. The cost of healthcare providers is increasing day by day. Thus, it is very important to predict the expected cost to avoid unexpected healthcare provider costs [1]. The cost utilization depends on the severity of the patients and their health conditions. Prediction in the early stage will be beneficial for the patient. In previous studies, ML techniques such as Support Vector Regressor (SVR) and Gradient Boosting Regressor (GBR) are used for the cost prediction of healthcare. However, the ML techniques do not perform accurately because the prediction results are heavily affected when the dataset is large, and contains the missing and outlier values [27]. A large volume of data is generated in the healthcare domain due to a massive number of patients. The data can be used for statistical analysis. The data is recorded in a non-proper way and not maintained properly according to the standards [1], [6]. The dataset comprising patients data mostly contains missing data, noisy data and inconsistent data that affect the performance of the model. The dataset also contains a massive number of attributes that do not impact the performance of the model. It consumes a lot of computational resources due to unnecessary information dissemination.

It also increases the time complexity of the algorithm that may overfit the model and affect prediction results [4]. The main critical challenge for prediction is that the classical prediction algorithms do not have generalization capability. The prediction does not always match with real world scenarios because the model is too complex and unable to generalize the solution with the best results. The performance of models is evaluated based on some defined performance metrics. It tells exactly about the model and checks whether the model makes accurate predictions or not. It simulates the positive and negative observations about the defined model. Choosing a limited number of performance metrics increases the chance of biasness or loss of information because the important performance metrics are not used for evaluation [22].

IV. PROPOSED SYSTEM MODEL

In the proposed model, the main contribution is the formation of hybrid models, SAE-DNN, VGG-DNN and VGG-SAE. The proposed models comprise of two stages: preprocessing stage and prediction stage. The focus of this study is to highlight the importance of prediction in the healthcare sector that helps healthcare providers to analyze their financial management. It aims to facilitate the patients to predict the expenditures of healthcare services to make financial decisions accordingly. The Healthcare Providers dataset is preprocessed using ML techniques. Standard Scaler is used for data scaling, Z-Score is used for removing the outliers and mode imputation is used for handling the missing values. After the preprocessing stage, the hyper-parameters of hybrid models are optimized using the random search algorithm. The first model in each hybrid combination is used to mitigate the dimensionality reduction problem in which the most important and relevant representation of feature vectors is extracted for accurate prediction. In VGG-DNN, VGG is used for dimensionality reduction. Hybrid DL model such as SAE-DNN is used for the cost prediction of healthcare providers, which helps the patients in healthcare utilization and making the decisions in healthcare according to their specific resources.

The Healthcare Providers dataset is available at Kaggle [37]. The dataset comprises comprehensive details regarding the financing of healthcare providers. It contains 100001 instances and 27 features. The important features selected are Average Medicare Standardized Amount, Average Medicare Payment Amount, Average Submitted Charge Amount, Average Medicare Allowed Amount, Number of Distinct Medicare Beneficiary/Per Day Services, Number of Medicare Beneficiaries and Number of Services. For the experimental purpose, the proposed approach is also validated on the Hospital Inpatient Cost Transparency: Beginning 2009 Dataset [38]. It provides detailed information regarding the 28 hospitals in New York and comprehensive details of healthcare providers. This dataset consists of 1048575 instances and 14 features. This dataset is collected from 2009 to 2016 but for the experimental testing, we only selected the most recent data, which is from 2015 and 2016. This section will provide the detailed insights into the proposed model. In Figure 1, the complete workflow of the proposed model is provided that covers all the components and their details.

A. DATA PREPROCESSING STAGE

The purpose of the data preprocessing stage is to remove unnecessary information from the dataset, which increases the performance of the model. The dataset contains some irrelevant information that increases the training time. However, it does not increase the performance of the model. This stage aims to extract only the useful features and represent them more effectively before training and testing the model to increase the accuracy and prediction of the specific model. Different ML techniques are employed to clean the data. Some of them are highlighted in this model such as mode imputation, Z-Score and standard scalar.

In handling the missing values, some entries in the dataset are found missing due to human error or incomplete information, which decreases the accuracy and performance of the model. Such values are denoted by NaN in the dataset. The issue must be encountered to overcome the overfitting issue caused due to the improper learning representation of the training data. So, mode imputation for continuous variables is used to replace the missing values with the mode value of the overall features instead of removing the missing values of the dataset. It also removes the crucial information from the dataset, which is very important for the prediction.

Without data normalization, values are scattered and unevenly distributed that leads to over-fitting, which affects the model's overall prediction. The model faces difficulty in learning the hidden patterns due to the scattered data points. For efficient and accurate cost prediction of healthcare providers, the feature vectors should be in a specific range. If the values are scattered, the model takes more time to train the feature vector of training data. Standard scaler normalization technique manages to scale the values between 0 and 1, and perform the linear transformation.

In the final step of preprocessing, the outliers are removed. This step plays a critical role in calculating the actual accuracy of the model. The outliers deviate from the actual range of values, which highly impacts the results of the prediction. Due to the presence of outliers in the dataset, the model learns the hidden patterns from the outlier's data, which results in poor training of the model and accurate prediction is not achievable. To overcome this issue, the Z-Score method is used to detect the outliers and remove them from the dataset for obtaining the best results of prediction.

B. PREDICTION STAGE

Cost prediction is always a concerning topic because it predicts the outcomes by mapping dependent and independent variables of the healthcare providers. After the data preprocessing stage, the next stage is prediction that estimates the cost expenditure of healthcare providers. Three different combinations of hybrid DL models, VGG-SAE, VGG-DNN and SAE-DNN, are used to optimize the outcomes of healthcare providers. In these combinations, SAE and VGG are used for dimensionality reduction. The model extracts the feature maps that are most relevant for the prediction that yields in result improvement of the hybrid model. Once the dimensions are reduced using different models, SAE-DNN, VGG-DNN and VGG-SAE are used for cost prediction of healthcare providers.

C. STACKED AUTOENCODER (SAE)

SAE is an unsupervised DL algorithm whose main objective is to select high-level representation and low-dimensional features from original input data using multiple layers of the Autoencoder (AE). AE is a deep neural algorithm that is used to extract low dimensional output. AE comprises a layered neural network that contains two components: encoder and decoder. The input data is passed to the encoder component that encodes the input data. It converts high-dimensional data to low-dimensional data by passing it through the bunch of hidden layers and by adjusting the weights of the input data. Once the input is compressed and converted into high-level representation without losing important information, the decoder component is utilized to reconstruct the input information. By utilizing the encoder and decoder components of AE, the dimensionality of original input data is reduced while retaining the crucial information [39].

Suppose the input features be x, the hidden layer y is given using Equation 1. The abbreviation and their definitions are shown in Table 2.

$$y^{(i)} = f(W_1^T x^{(i)} + b_1) \tag{1}$$

where $f = tanh(.)$ is an activation function.

$$Z^{(i)} = W_2^T y^{(i)} + b_2 \approx x^{(i)} \tag{2}$$

The AE will be trained by reducing the objective function [39]

$$J(X, Z) = \frac{1}{2} \sum_{i=1}^{m} \left\| y^{(i)} - z^{(i)} \right\|^2 \tag{3}$$

where m denotes the feature vectors data points.

TABLE 2. Abbreviation and definitions.

Abbreviation	Definitions
b_n	Bias vectors
$J(X, Z)$	Objective function
k	Hidden layers
$tanh$	Activation function
W_n	Weights
X	Input data
Y_k	Output of hidden layers
$\hat{y}$	Predicted value
$Z^{(i)}$	Output of decoder

SAE is a multi-layered AE model in which multiple AEs are stacked on top of one another to extract the low-dimensional features. It enhances the performance of SAE. It trains the model using the original input data and passes the outcomes of previous layer as input to the proceeding layer unless training is finished. It utilizes the back propagation algorithm to minimize the cost function. It updates the weights for fine-tuning parameters. The input data is passed to the first AE. Output of the first AE and the input of first AE are fed to the second AE. Output and input of the second AE are fed to the third AE. Through this process, compression and decompression are performed on every AE and SAE algorithm is executed layer by layer. Using more than one AE, SAE performs the high-level representation of original features more perfectly than individual AE [40], [41]. The main advantage of SAE is that it efficiently handles the complex relationship within the features of datasets. Due to more AE layers in SAE, it performs dimensionality reduction without losing the important features. SAE decreases the amount of noisy values from the training data. It also decreases the computational time and the resource consumption due to the conversion of high-dimensional data into low-dimensional data.

The dimensionality of input data is reduced that results in minimizing the computational time because the irrelevant information is discarded. The working of SAE is given in Algorithm 1.

Algorithm 1 Stacked Autoencoder

Input: Input set $X = \left\{x^{(i)}, 1 \leq i \leq m, x^i \epsilon R^n\right\}$, the number of hidden layers k

Output: Updated outputs' weights $\{W_h, 1 \leq h \leq k\}$, bias vectors' values $\{b_h, 1 \leq h \leq k\}$, the output Y_k of the k-th hidden layer

1: $Y_0 = X$;
2: W_1 of the first hidden layer are trained with X, and results of Y_1 are obtained with W_1
3: The output of previous layer Y_1 is passed as an input to the next AE, obtaining W_2 and Y_2
4: The algorithm will be repeated till the k-th hidden layer, and results in W_k and Y_k
5: SAE arranges all the trained AE and performs fine-tuning of the weights and biases using the Back propagation.
6: Finally, the weights $\{W_h, 1 \leq h \leq k\}$, bias vectors $\{b_h, 1 \leq h \leq k\}$, and ${Y_k}^i$ $(1 \leq i \leq m)$ are obtained.

D. DEEP NEURAL NETWORK (DNN)

DNN is a feed-forward neural network that utilizes more than two hidden representations to generate optimized results. The training data is passed to the input layer, then to the hidden layers and then to the output layer, which is responsible to produce results. DNN contains the input layer, at least two hidden layers, and an output layer [42]. The basic difference between ANN and DNN is that ANN can work

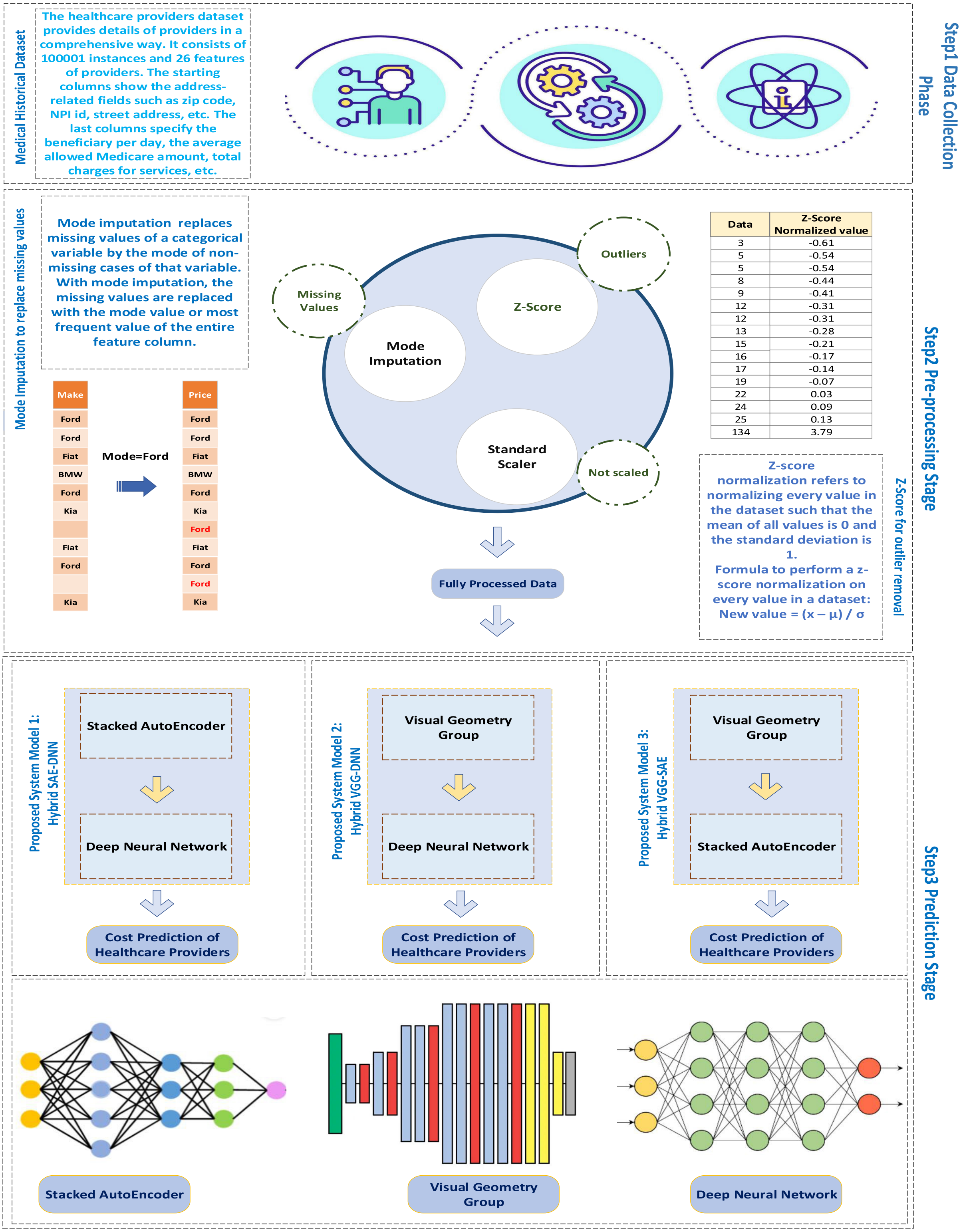


FIGURE 1. Proposed system model.

with one hidden layer as well. However, DNN must have at least two hidden layers. DNN utilizes its hidden layers to learn the hidden patterns by mapping the dependent and the independent variables [43].

The hidden layers of DNN represent a fully connected network, which exhibits that any neuron from the previous layer is connected to every neuron in the next layers and learns the hidden pattern from the training data. The main advantage of DNN is that it easily handles labeled and unlabeled data. Due to multiple hidden layers, DNN can easily learn the complex relationships from large amounts of training data. The DNN model comprises of a linear activation function given in Equation 4.

$$a = \sum w_i x_i + b_i \tag{4}$$

where the inputs of each neuron are represented by x_i; w_i represents the weight and b_i represents bias.

Multiple hidden layer's outputs can be expressed as follows.

$$\begin{aligned} f(x) &= f[a^{L+1}(h^L(a^L(\cdots(h^2(a^2(h^1(a^1(x))))))))] \\ a^L(x) &= W^L x + b \end{aligned} \tag{5}$$

where the hidden layers are denoted by L and f(x) shows the output of DNN [44], [45]. The working of DNN is given in Algorithm 2.

Algorithm 2 Deep Neural Network

Input: X,Y of dataset with size $[N, d_x]$ and $[N, 1]$, respectively.

Output: updated weights w_i, bias values b_i

1: $w_0 = Parameter\ (size = [d_0, d_x])$
2: $h_0 = ReLU\left(Xw_0^T\right)[N, d_0]$
3: $w_1 = Parameter\ (size = [d_1, d_0])$
4: $h_1 = ReLU\left(h_0 w_1^T\right)[N, d_1]$
5: $w_2 = Parameter\ (size = [1, d_1])$
6: $\hat{y} = ReLU\left(h_1 w_2^T\right)[N, 1]$
7: $l = \left\|\hat{y} - \mathbf{y}\right\|^2 LossFunction$
8: $\mathbf{w} = (w_0, w_1, w_2)$
9: **repeat**
10: $\mathbf{w} = \mathbf{w} - \rho \nabla_w l$
11: **until** convergence

E. VISUAL GEOMETRY GROUP (VGG-11)

VGG was proposed by Karen Simoyan and Andrew Zisserman of Oxford University in 2014. VGG-11 means that it has 11 layers. It is formulated to understand the depth of CNN by increasing the hidden layers and reducing the kernel size. It aims to reduce the number of parameters in the convolution layer by introducing the fixed size kernel filter. VGG is the variant of CNN that uses the convolution concepts, which make the model best fit for the time series data. CNN comprises of three main components: convolution layer, pooling layer, and fully connected layer. The input sequence is first passed to the convolution layer, which convolves the input data in such a way that it extracts the important features from the input. This layer uses the filter map concept, which formulates the output of the convolutional layer by iterating through the input sequence. It calculates the filter values by taking the dot product of filter values and input values. The output of the convolutional layer is dimensionally high. The pooling layer is utilized after the convolutional layer, which receives the feature vector of the convolutional layer and minimizes the dimensionality of the extracted features. Then, the output is passed to the fully connected layer, which takes the extracted features as input and is responsible for the final output of the model as it follows the neural network structure to move the input sequence from the multiple hidden layers. The activation function is utilized to obtain the desired output [46], [47].

Each convolutional layer comprises a multiple number of kernels and their calculations are performed by following equation:

$$l_t = tanh\,(x_t * k_t + b_t) \tag{6}$$

where l_t represents the output of convolution, the activation function is tanh, x_t shows the input vector, k_t is the weight of the convolution layer, and b_t is the bias [46].

The pooling layer is utilized to minimize the high dimensionality of extracted vectors by estimating the value of region and calculated by the following equation:

$$a_j^{l+1}(c, d) = \max_{0 \le p,q < m} \left\{ a_j^l\,(c \cdot m + p, d \cdot m + q) \right\} \tag{7}$$

The fully connected layer is utilized after the pooling layer. The features are extracted and dimensionality is reduced. It is used to train the model and to obtain the desired outputs [48].

VGG also has the same structure as that of CNN and comprises of convolutional layer, pooling layer and fully connected layer. The difference is in the number of convolutional layers, change in parameters of the convolutional layer, specific selection of approach in the pooling layer, the change in configuration of pooling layer, increase in the number of fully connected layers and parametric values of the fully connected layer.

Convolutional Layer: At this layer, the kernel of size 3*3, stride with value 1 and ReLU activation function are used. VGG-11 consists of 8 convolutional layers.

Max Pooling Layer: At this layer, max pooling is used that selects the maximum value from the extracted feature map of the convolutional layer. Stride is fixed to 1 that defines the procedural steps for the hidden layers.

Fully Connected Layer: The output of max pooling layer is passed to the fully connected layer that is responsible for the training of extracted features. Three fully connected layers are used that are responsible to transform the output of the model. In the proposed model, 1D convolutional layer with kernel size 3 and 1D pooling layer are utilized because we are dealing with the time series data [49]. The main advantage of VGG is better prediction results due to the increase in the number of convolutional layers. VGG improves the depth of the model using the small kernel values. The working of VGG-11 is given in Algorithm 3.

Algorithm 3 Visual Geometry Group

Input: Input data, $X = \{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\}$, Y of dataset.

Output: Updated weights, bias vectors and output of hidden layer.

$C_1 = Parameter\ (size = [64, MaxP])$
$C_2 = C_1\ (size = [128, MaxP])$
$C_3 = C_2\ (size = [256])$
$C_4 = C_3\ (size = [256, MaxP])$
$C_5 = C_4\ (size = [512])$
$C_6 = C_5\ (size = [512, MaxP])$
$C_7 = C_6\ (size = [512])$
$C_8 = C_7\ (size = [512, MaxP])$
$FC1 \leftarrow FC\ (C_8, size = [4096])$; where $FC()$ represents fully-connected layer for regression.
$FC2 \leftarrow FC1\ (size = [4096])$
$FC3 \leftarrow FC2\ (size = [1000])$
$\mathbf{Y} = FC3$

F. SUPPORT VECTOR REGRESSOR (SVR)

SVR is a supervised ML technique that can easily be used for linear and non-linear data points. SVR is purely based on the SVM for the cost prediction of healthcare. SVR classifies and divides the data points into different classes. The separation of data points based on the hyperplane depends on the nature of the dataset. So, the SVR finds the best decision boundary that optimally separates the data points. The data points nearest to the hyperplane are known as the support vectors. SVM uses the support vector points to create a hyperplane and maximizes the margin between the line and the support vector points [50]. The hyperplane can be represented using Equation 8.

$$y = wX + b \tag{8}$$

where X represents the data points of the dataset, w represents the weights and b highlights the intercept at zero value.

SVR is primarily used for the regression task. The main advantages of SVR are that it performs better when the data is linear and high dimensional. Due to the utilization of hyperplane and support vector points, SVR is also found to be memory efficient.

G. GRADIENT BOOSTING REGRESSOR (GBR)

GBR is an ML technique that aims to convert weak learners into strong learners. It creates an ensemble by combining multiple weak learners. It uses multiple same-sized decision trees and adopts the iterative approach in which every iteration strengthens the ensemble model. Moreover, GBR is considered to be the best fit for finding the complex relationships between the features and the independent variables of the dataset [51]. To optimize the GBR prediction, the prediction error needs to be minimized, which is done using Equation 9.

$$r_1 = y - mean\ (y) \tag{9}$$

where r_1 represents the residual errors of prediction and y denotes the target variable. The main advantage of GBR is the efficient handling of missing values and outliers. It easily computes the non-linear relationship among the features. It can be easily trained on different loss functions and performs better with numerical and categorical values.

Algorithm 4 Working of Proposed Model 1: Hybrid SAE-DNN

Input: $X = \{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\}$, Y of dataset
Output: Output Y of hybrid SAE-DNN model

Take X and Y from Healthcare Providers Dataset
Apply Mode imputation on X and Y
Apply Z-Score on X and Y
Apply Standard Scaler Normalization on X and Y
Split the X and Y into X_{train}, Y_{train}, X_{test} and Y_{test}
SAE working mechanism:
Input layer $S_{\text{i}} = X_{\text{train}}$ and Y_{train}
Encoder layer $E_{\text{i}} = S_{\text{i}}$
Decoder layer $D_{\text{i}} = E_{\text{i}}$
Train the SAE using encoder and decoder
Pass the output of D_{i} to DNN
DNN working mechanism:
Input layer $y_{\text{i}} = D_{\text{i}}$
Hidden layer $h_i = ReLU\ \left(D_{\text{i}} w_0^T\right)$
$w_i = Parameter\ (size = [d_0, d_x])$
Output layer $\hat{y} = relu\ \left(h_1 w_2^T\right)\ [N, 1]$
$l = \left\|\hat{y} - \mathbf{y}\right\|^2\ LossFunction$
Performance Metrics
$MSE \leftarrow meanSquaredError(Y_{\text{test}}, y_{\text{pred}})$
$\text{RMSE} \leftarrow \sqrt{\text{mse}}$
$\text{MAE} \leftarrow meanAbsoluteError(Y_{\text{test}}, y_{\text{pred}})$
$R - squared \leftarrow R - squaredscore(Y_{\text{test}}, y_{\text{pred}})$

In the hybrid SAE-DNN model, input features from the healthcare providers data are passed to the SAE-DNN and the cost of the healthcare providers is estimated. In DL models, the values of hyper-parameters play an important role in the training of data as the best-fit learning of the model and performance depend on these values. The hyper-parameters of SAE-DNN are epoch, batch size, activation function, optimizer, and loss function. Number of epochs is selected to be 20 and the batch size is taken as 32. Moreover, ReLU activation function, Adam optimizer and MSE, RMSE, MAE, and R-squared, as a loss metric are used for SAE-DNN. After the sequence of SAE, DNN is used to predict the cost estimation of healthcare providers. DNN consists of multiple hidden layers and dense layers that perform back propagation to optimize the weights, which helps in efficient prediction of healthcare providers. The working of hybrid SAE-DNN is given in Algorithm 4 and the hyper-parameters of SAE-DNN are provided in Table 3.

In the hybrid VGG-DNN model, the Healthcare Providers' Dataset is cleaned using the ML techniques. After the preprocessing of the dataset, scaled features of the input data

TABLE 3. Hyper-parameters' values of proposed model 1: hybrid SAE-DNN.

Case	Hyper-parameter	Values
1	Epoch	20
2	Batch Size	32
3	Optimizer	Adam
4	Loss	MSE
5	Activation	ReLU, Linear

Algorithm 5 Working of Proposed Model 2: Hybrid VGG-DNN

Input: Input data as $X = \left\{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\right\}$, Y of dataset

Output: The output Y of hybrid SAE-DNN model

Take X and Y from Healthcare Providers Dataset
Apply Mode imputation on X and Y
Apply Z-Score on X and Y
Apply Standard Scaler Normalization on X and Y
Split the X and Y into X_{train}, Y_{train}, X_{test} and Y_{test}
VGG working mechanism:
Input layer $S_{\text{i}} = X_{\text{train}}$ and Y_{train}
$C_i = C_i\,(size = [Filter_{\text{i}}, MaxP])$
$FC_i \leftarrow FC_i\,(C_i, size = [Dense_i])$; where $FC()$ represents fully-connected layer for regression
Train the VGG using multiple C_i and FC_i
DNN working mechanism:
Input layer $y_{\text{i}} = D_{\text{i}}$
Hidden layer $h_i = ReLU\left(D_{\text{i}} w_0^T\right)$
$w_i = Parameter\,(size = [d_0, d_x])$
Output layer $\hat{y} = ReLU\left(h_1 w_2^T\right)[N, 1]$
$y_{pred} = \left\|\hat{y} - \mathbf{y}\right\|^2 LossFunction$
Hybrid working mechanism:
concatenate the output of VGG and DNN
$Concat = \left(VGG_{output}, DNN_{output}\right)$
Train the hybrid model using the VGG and DNN
Performance Metrics
$MSE \leftarrow meanSquaredError(Y_{\text{test}}, y_{\text{pred}})$
$\text{RMSE} \leftarrow \sqrt{\text{mse}}$
$\text{MAE} \leftarrow meanAbsoluteError(Y_{\text{test}}, y_{\text{pred}})$
$R - squared \leftarrow R - squaredscore(Y_{\text{test}}, y_{\text{pred}})$

are fed to the VGG-11 that is used to perform dimensionality reduction of the input data. It consists of eight convolutional layers and three dense layers that aim to explicitly extract the useful features and reduce the dimensionality of the obtained features. The extracted features are passed to the DNN model that accepts the feature and trains the model on training data. It learns the hidden pattern from the dataset. The hidden layers and the dense layer of DNN are used to finally predict the outcomes of cost regarding the healthcare providers. The hyper-parameters' values of VGG-DNN are fundamental for the prediction. The results depend on the values of hyper-parameters. The number of epochs is taken to be 20 with the batch size 32. Also, Adam optimizer is used along with ReLU activation function. The loss functions used such as MSE, RMSE, MAE and R-squared. While, the value of both linear and padding value is the same. The complete details of VGG-DNN are provided in Algorithm 5 and hyper-parameters are well defined in Table 4.

TABLE 4. Hyper-parameters' values of proposed model 2: hybrid VGG-DNN.

Case	Hyper-parameter	Values
1	Epoch	20
2	Batch Size	32
3	Optimizer	Adam
4	Loss	MSE
5	Activation	ReLU, Linear, Softmax

Algorithm 6 Working of Proposed Model 3: Hybrid VGG-SAE

Input: $X = \left\{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\right\}$, Y of dataset

Output: Weights, bias vectors and the output Y of hybrid SAE-DNN model

Take X and Y from Healthcare Providers Dataset
Apply Mode imputation on X and Y
Apply Z-Score on X and Y
Apply Standard Scaler Normalization on X and Y
Split the X and Y into X_{train}, Y_{train}, X_{test} and Y_{test}
VGG working mechanism:
Input layer $S_{\text{i}} = X_{\text{train}}$ and Y_{train}
$C_i = C_i\,(size = [Filter_{\text{i}}, MaxP])$
$FC_i \leftarrow FC_i\,(C_i, size = [Dense_i])$; where $FC()$ represents fully-connected layer for regression
Train the VGG using multiple C_i and FC_i
SAE working mechanism:
Input layer $S_{\text{i}} = X_{\text{train}}$ and Y_{train}
Encoder layer $E_{\text{i}} = S_{\text{i}}$
Decoder layer $D_{\text{i}} = E_{\text{i}}$
Train the SAE using encoder and decoder
Hybrid working mechanism:
concatenate the output of VGG and SAE
$Concat = \left(VGG_{output}, SAE_{output}\right)$
Train the hybrid model using the VGG and SAE
Performance Metrics
$MSE \leftarrow meanSquaredError(Y_{\text{test}}, y_{\text{pred}})$
$\text{RMSE} \leftarrow \sqrt{\text{mse}}$
$\text{MAE} \leftarrow meanAbsoluteError(Y_{\text{test}}, y_{\text{pred}})$
$R - squared \leftarrow R - squaredscore(Y_{\text{test}}, y_{\text{pred}})$

In the hybrid VGG-SAE model, input data is first preprocessed. Then, the VGG is used for dimensionality reduction and SAE is used for cost prediction of healthcare providers. The hyper-parameters' values of VGG-DNN are fundamental for the prediction as the results depend on them. The number of epoch is 20, the batch size is 32, the optimizer is Adam, the loss functions such as MSE, RMSE, MAE and

TABLE 5. Hyper-parameters' Values of Proposed Model 3: Hybrid VGG-SAE.

Case	Hyper-parameter	Values
1	Epoch	20
2	Batch Size	32
3	Optimizer	Adam
4	Loss	MSE
4	Activation	ReLU, Linear, Softmax
5	Padding	Same

R-squared, and the activation functions are ReLU and Linear. VGG uses convolution and max pooling to extract the features more accurately than traditional ML techniques and reduce the dimensions. DNN also uses its hidden layers to accurately predict the cost of healthcare providers. By VGG-SAE, the prediction results are improved as compared to individual VGG and SAE. The overview of the proposed VGG-SAE is given in Algorithm 6 and the details of hyper-parameters are provided in Table 5.

H. MODEL INTERPRETATION

DL models exhibit the black box approach where the input is passed to each model and the respective output is generated. The internal working of the models is unknown and how the models calculate the output is also hidden. SHAPley is used to explain the output of the hybrid models. The output of each hybrid models is passed to SHAPley as an input. It further breakdowns the output and explains the contributions of each model. It also explicitly explains the importance of each feature in the prediction outcome. It uses the coalition game in which multiple players are selected. Each player is awarded based on its performance. In SHAPley, SHAP values and explainers are used to highlight the importance of each feature in the output. SHAP values are represented as a matrix that exhibits the score of each feature [52]. Explainers receive the desired model as input and extract the importance of each feature in the prediction using Equation 10.

$$g\left(z'\right) = o_0 + \sum_{j=1}^{M} \o_j z'_j \tag{10}$$

where z' represents the features of hybrid model, M represents the total size of features and o explains the score values for each vector.

V. SIMULATION RESULTS

The simulation results and their discussions are elaborated in this section. All combinations of DL models are executed three times and the best values are shown in the simulation results. Simulation results are obtained using a COLAB notebook. Python language is used to implement the proposed models. The results are recorded over the same time frame. The results are obtained using multiple DL models such as SAE, DNN and VGG, and their different combinations such as VGG-SAE, VGG-DNN and SAE-DNN. To understand the contribution of each feature in the prediction results, SHAPley is used on VGG-SAE, VGG-DNN and SAE-DNN. The SHAPley model is evaluated on a 50% dataset to explain the score of each feature. The Healthcare Providers Dataset is evaluated with and without using the optimized parameters. Without hyper-parameter tuning, the hyper-parameters of these DL models are the same for each individual and hybrid combination. The epoch and batch size are the most important hyper-parameters of any DL model used for both single and hybrid models. The number of epochs is taken as 20 while the batch size is taken as 32. In addition, Adam optimizer is used along with ReLU, linear and softmax as activation functions, and MSE as a loss metric for all single and hybrid models used in this paper. The number of epoch is selected to be 20 because all the models stop approximately in between 18 to 23 epoch. So, our models perform the best in this range and if the epoch size increases to 40 or 50, the model will overfit and the prediction results will be unfavorable. So, 20 is the best value of epoch for this specific problem. Secondly, the standard rule of thumb regarding the values of batch size is 32 in which the model performs the best using these hyper-parameters. To evaluate the performance of these DL models, different performance metrics are used such as MSE, RMSE, MAE, R-squared and the execution time. With hyper-parameter tuning on Healthcare Providers Dataset, different hyper-parameter values are obtained according to the nature and complexity of the hybrid models. For VGG-SAE, the epoch value is 20, the learning rate is 0.001, and the batch size is 32. For VGG-DNN, the learning rate is 0.01, the epoch value is 25, and the batch size is 64. For SAE-DNN, the learning rate is 0.01, the epoch value is 30, and the batch size is 16. The random search technique is also implemented on Hospital Inpatient Cost Transparency dataset and different hyper-parameter values are obtained. For VGG-SAE, the epoch value is 25, the learning rate is 0.01, and the batch size is 32. For VGG-DNN, the learning rate is 0.1, the epoch value is 20, and the batch size is 32. For SAE-DNN, the learning rate is 0.01, the epoch value is 30, and the batch size is 16.

The results of these DL models are explicitly mentioned in Tables 6, 7, 8 and 9. According to Table 6, individual models such as SAE, DNN, and VGG take less time than hybrid models. VGG takes more time to train the model because VGG is an extensive network and is trained on approximately 138 million parameters that results in better performance. However, VGG takes more time to train its parameters on the training data. In Table 7, the results are obtained using the Hospital Inpatient Cost Transparency dataset in the absence of optimized hyper-parameters. Different evaluation metrics are used to analyze the performance of the DL models. The DL models also perform better on the Hospital Inpatient Cost Transparency dataset. The hybrid models such as VGG-SAE, VGG-DNN, and SAE-DNN perform better than their individual models such as VGG, DNN, and SAE. In Tables 8 and 9, the result of performance metrics are provided when using the optimized parameters for two different datasets. The results are better than the results obtained without using the optimized parameters. The proposed hybrid DL models such as VGG-SAE, VGG-DNN, and SAE-DNN are compared with the ML techniques such as SVR and GBR. The hybrid models are superior to ML techniques because the DL models more efficiently predict the cost of healthcare providers than ML techniques. In Tables 6 and 7, the prediction results of DL models and ML techniques without using the optimized hyper-parameters are elaborated. The MSE, RMSE, MAE, and R-squared errors of ML techniques are more than the errors obtained using the proposed DL models, which confirms that the proposed models are better and more accurate than ML techniques.

Different evaluation parameters are used to determine the performance of each model. MSE also plays an important role in finalizing the best model. It is the average squared difference between the actual and the predicted values. MSE shows how well the model predicts the data. In Figures 2 and 3, the MSE values are depicted for each DL model. In the context of MSE values, the hybrid models such as VGG-SAE and SAE-DNN exhibit the same behavior. VGG-SAE and SAE-DNN have the same MSE value that is 0.01. In the hybrid models, the MSE of VGG-DNN is increased by 0.1 than other deep models such as VGG-SAE and SAE-DNN. However, it also competes with its individual models such as VGG and DNN. This value indicates that the proposed hybrid models will predict the data accurately and best fit the training data. The hybrid model takes more time than individual techniques. However, in the case of MSE, SAE-DNN, VGG-SAE and VGG-DNN outperform the standard SAE, DNN and VGG. DNN performs the worst and its value of MSE is the highest that is 0.98 because DNN tends to face problems in interpreting the results.

RMSE is the square root of the MSE between the predicted and the actual values [53]. In Figures 4 and 5, the RMSE values are explicitly shown against the deep models. DNN performs the worst in the context of RMSE. It has the highest RMSE value, i.e., 0.99, which means it does not fit the model correctly and does not predict the actual values accurately. SAE-DNN also performs the best as it beats the SAE and DNN with RMSE value of 0.12. In SAE-DNN, SAE is responsible for feature engineering, which aims to transform the input features into high-level representations of the extracted features. The transformation helps the model to map the feature vectors perfectly, which decreases the error between the predicted and the actual values. Multiple hidden layers with small and fixed values of filter maps help the VGG combinations to map the features more precisely than DL models and improve the prediction error. The RMSE of SAE is 0.44 and that of DNN is 0.99, so the SAE-DNN, VGG-DNN, and VGG-SAE outperform the individual models in terms of RMSE.

In Figures 6 and 7, MAE values are calculated for each deep model. DNN performs the worst in the context of MAE. The MAE value of DNN is 0.24 that is the highest of all deep models. SAE and VGG exhibit better performance in the context of MAE value. The hybrid models such as VGG-SAE and VGG-DNN exhibit the same MAE value that is 0.03. The value of 0.03 shows that the hybrid models outperform the individual models and perform perfect mapping between the independent and the dependent variables. SAE-DNN also outperforms its baseline models. The MAE value of SAE is 0.03 and the MAE value of DNN is 0.24. VGG-DNN also performs better than VGG and DNN as the error in VGG-DNN is less than both VGG and DNN, which is 0.19 and 0.24, respectively. However, the MAE value of VGG-DNN is 0.03. The third combination, which is VGG-SAE, also performs better than individual models, VGG and SAE. VGG-SAE performs better than VGG and SAE. VGG takes more training time to train its parameters that are approximately 138 million in number. It tends to reduce the dimensions of the features using its pooling layer. However, it increases the overall performance of the prediction model. In the SAE-DNN model, SAE is used to refine the input in such a way that the input features are fully transformed into a high representation of extracted features. It helps the DNN model to pass only the important features for training the model. SAE-DNN, VGG-DNN, and VGG-SAE are outperforming combinations of models, which make these deep models better than VGG, SAE, and DNN.

R-squared determines the best fit of models and indicates the percentage of the variance between the dependent and independent variables. Its value lies between 0 and 1. If the values are close to 1, it means that the model performs perfectly and correlate. In Figures 8 and 9, R-squared values are plotted against their models. The hybrid models perform the best according to their R-squared value of 0.98, which indicates that the model best fits on the training data. DNN exhibits the least R-squared value of 0.16, which means that the DNN model does not best fit the model and does not perform accurately. DNN contains at least two hidden layers that lack in learning hidden patterns from the extracted features and poorly map the correlation between dependent and independent feature vectors. SAE-DNN outperforms SAE and DNN because the encoder and decoder functionality of SAE reduces the dimensions of the provided input features that results in high predictive performance and less time taken by the model. DNN performs the worst because it is computationally expensive and has more chance of over-fitting, which lessens the capability of learning the hidden features. The SAE-DNN also outperforms its baseline models, which are SAE and DNN. The R-squared value for SAE-DNN is 0.98, for SAE is 0.82 and for DNN is 0.16. VGG-DNN also performs better than its baseline models. The R-squared value of VGG-DNN is 0.97, for VGG is 0.69 and for SAE is 0.82. VGG-SAE also outperforms VGG. The VGG model is trained on 138 million parameters that results in increase in the performance of the model. The R-squared value of SAE-DNN is 0.98, whereas the R-squared value of

TABLE 6. Prediction results with hybrid deep learning models for the healthcare providers dataset without using optimized hyper-parameters.

Performance metrics	SAE	DNN	VGG	VGG-SAE	VGG-DNN	SAE-DNN	SVR	GBR
MSE	0.19	0.98	0.35	0.01	0.02	0.01	0.64	0.19
RMSE	0.44	0.99	0.57	0.13	0.16	0.12	0.80	0.43
MAE	0.03	0.24	0.19	0.03	0.03	0.02	0.06	0.13
R-squared	0.82	0.16	0.69	0.97	0.97	0.98	0.45	0.83
Execution Time (s)	251	202	645	751	584	450	240	45

TABLE 7. Prediction results with hybrid deep learning models for the hospital inpatient cost transparency dataset without using optimized hyper-parameters.

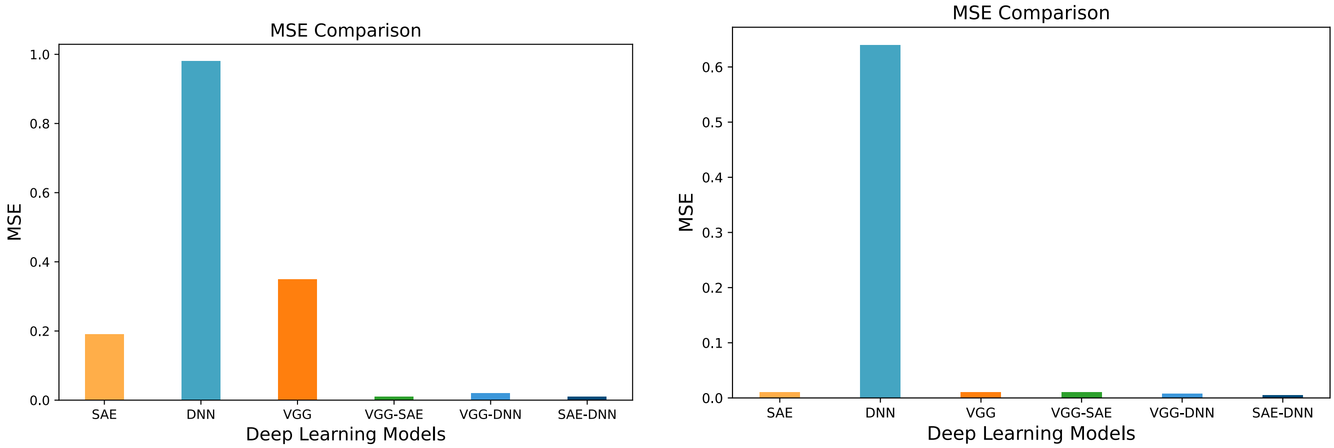
Performance metrics	SAE	DNN	VGG	VGG-SAE	VGG-DNN	SAE-DNN	SVR	GBR
MSE	0.01	0.64	0.01	0.01	0.008	0.005	0.11	0.20
RMSE	0.13	0.80	0.13	0.13	0.09	0.07	0.33	0.44
MAE	0.06	0.36	0.07	0.05	0.03	0.03	0.05	0.21
R-squared	0.98	0.33	0.98	0.98	0.99	0.99	0.88	0.79
Execution Time (s)	500	380	1340	1640	620	818	450	80

TABLE 8. Prediction results with optimized parameters on hybrid deep learning models for the healthcare providers dataset using optimized hyper-parameters.

Performance metrics	VGG-SAE	VGG-DNN	SAE-DNN
MSE	0.01	0.01	0.01
RMSE	0.13	0.12	0.11
MAE	0.02	0.02	0.02
R-squared	0.98	0.99	0.99
Execution Time (s)	780	548	530

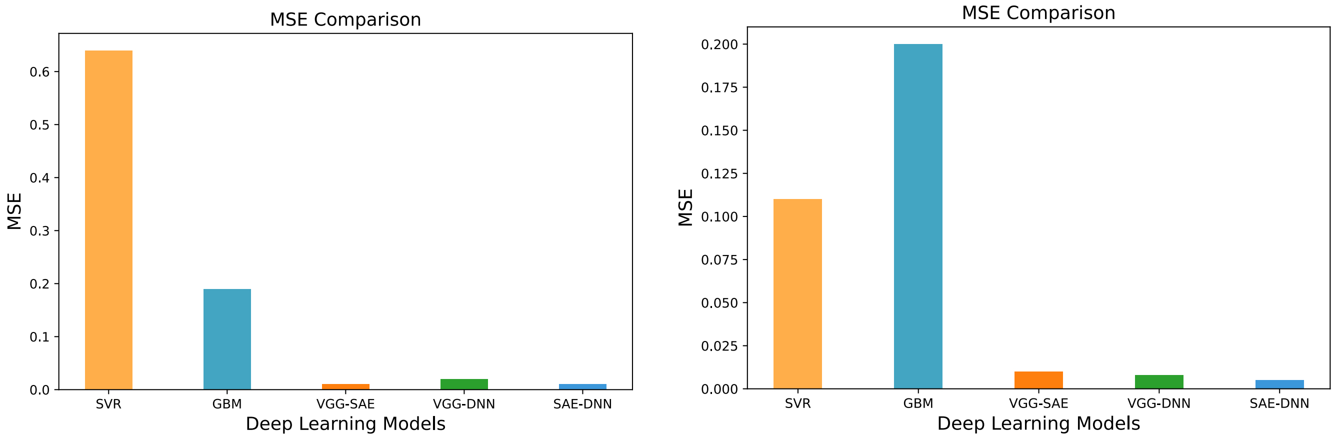
TABLE 9. Prediction results with optimized parameters on hybrid deep learning models for the hospital inpatient cost transparency dataset using optimized hyper-parameters.

Performance metrics	VGG-SAE	VGG-DNN	SAE-DNN
MSE	0.007	0.006	0.003
RMSE	0.08	0.08	0.06
MAE	0.03	0.03	0.02
R-squared	0.99	0.99	0.99
Execution Time (s)	1680	645	850



(a) MSE Plot for the Healthcare Providers Dataset (b) MSE Plot for the Hospital Inpatient Cost Transparency Dataset

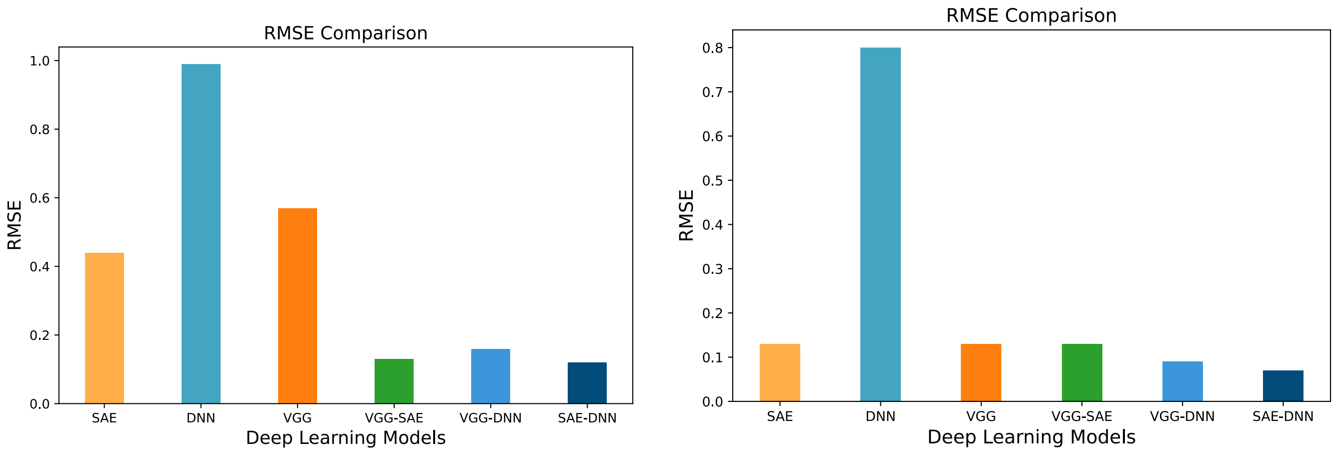
FIGURE 2. MSE plots of individual and hybrid DL models for the healthcare providers and hospital inpatient cost transparency datasets.



(a) MSE Plot for the Healthcare Providers Dataset

(b) MSE Plot for the Hospital Inpatient Cost Transparency Dataset

FIGURE 3. **MSE comparison of ML techniques with hybrid models on healthcare providers and hospital inpatient cost transparency datasets.**



(a) RMSE Plot for the Healthcare Providers Dataset

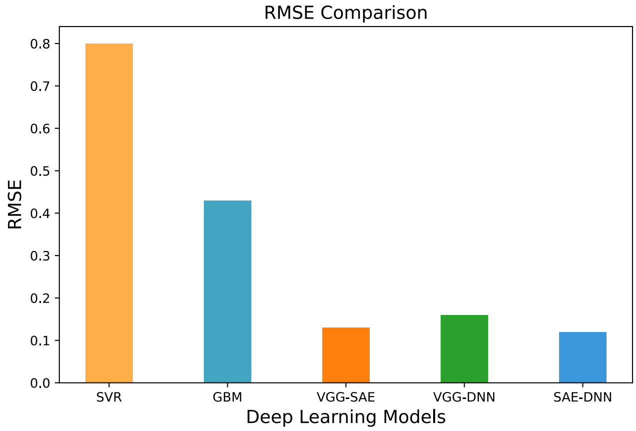
(b) RMSE Plot for the Hospital Inpatient Cost Transparency Dataset

FIGURE 4. **RMSE comparison of individual and hybrid DL models for the healthcare providers and hospital inpatient cost transparency datasets.**

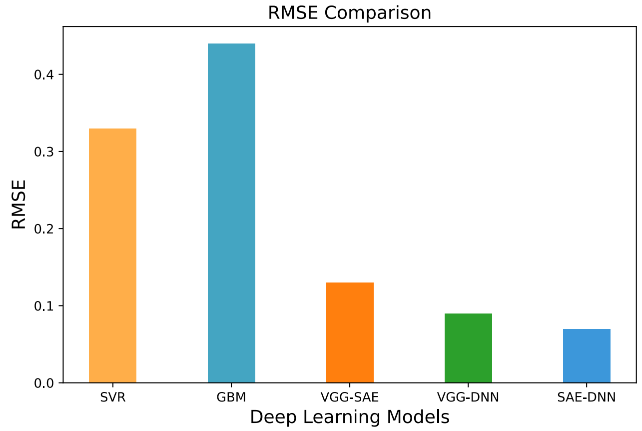
VGG is 0.69. While it is 0.82 for SAE. So, SAE-DNN, VGG-DNN, and VGG-SAE are the best fit for the training data.

Computational time is the time taken by the model training in which the model learns the hidden patterns. It maps the relationship between the dependent and independent variables of the training data. In Figures 10 and 11, the execution time is explicitly shown for each DL model. VGG-SAE takes the maximum time among all the hybrid models that is 12 minutes and 31 seconds. It is because VGG contains eight convolutional layers. Each convolutional layer contains fixed kernel values. However, DNN takes the minimum time, which is 3 minutes and 22 seconds. It is because DNN requires at least two hidden layers to train the model that consumes less time. However, the model gives low performance. By considering all the hybrid combinations of DL models, SAE-DNN takes the least time that is 7 minutes and 30 seconds whereas VGG-DNN takes 9 minutes and 43 seconds. By looking at Figure 10, from all the individual models, VGG takes the maximum time that is 10 minutes and 45 seconds while SAE takes less time that is 4 minutes and 11 seconds. The hybrid models take more time than individual models. However, they increase the prediction power of the model by correctly predicting the values.

The optimal selection of features is validated using the SHAPley model. SHAPley explicitly provides the importance of each feature of the dataset. Different plots such as summary plot, bar plot, dependence plot and force plot are used for VGG-SAE, VGG-DNN and SAE-DNN. In Figures 12 and 13, the dependency plots of VGG-SAE, VGG-DNN and SAE-DNN model for Average Submitted Charge Amount and Average Medicare Payment Amount features are shown. In the dependency plots, the SHAP value increases with the feature value which means that the features have a positive effect on the prediction outcomes of VGG-SAE, VGG-DNN and SAE-DNN. In Figures 14, 15 and 16, the results of the SHAP model on the VGG-SAE model

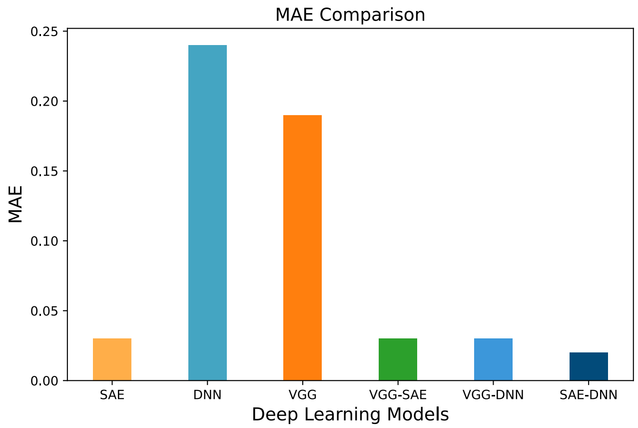


(a) RMSE Plot for the Healthcare Providers Dataset

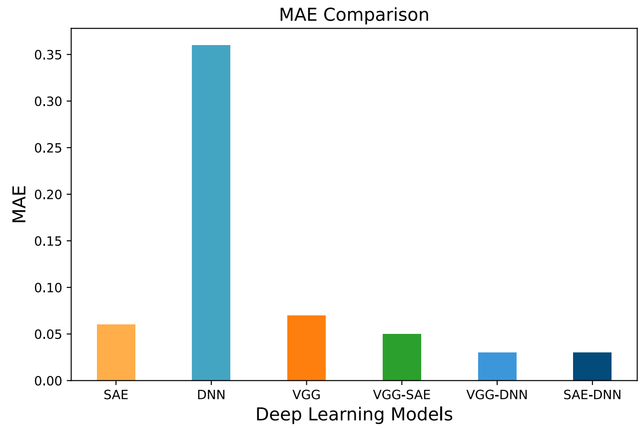


(b) RMSE Plot for the Hospital Inpatient Cost Transparency Dataset

FIGURE 5. **RMSE comparison of ML techniques with hybrid models on healthcare providers and hospital inpatient cost transparency datasets.**

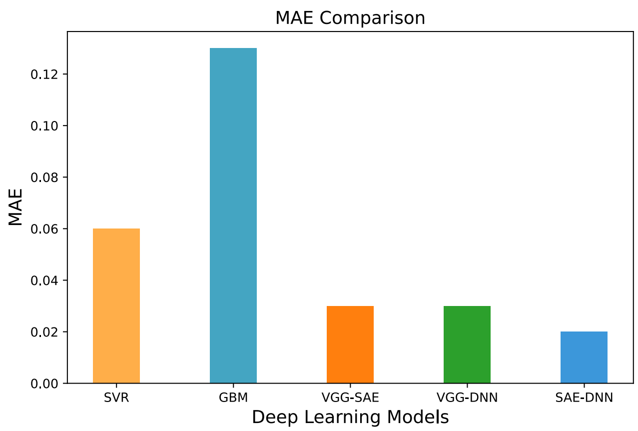


(a) MAE Plot for the Healthcare Providers Dataset

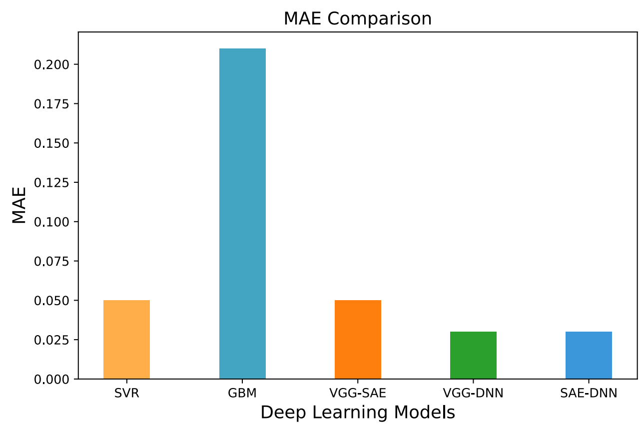


(b) MAE Plot for the Hospital Inpatient Cost Transparency Dataset

FIGURE 6. **MAE comparison of individual and hybrid DL models for the healthcare providers and hospital inpatient cost transparency datasets.**

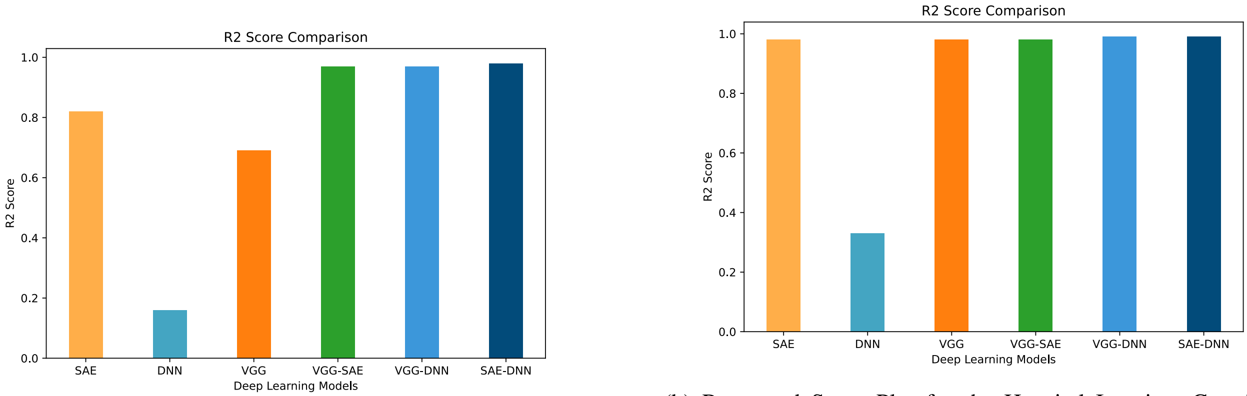


(a) MAE Plot for the Healthcare Providers Dataset



(b) MAE Plot for the Hospital Inpatient Cost Transparency Dataset

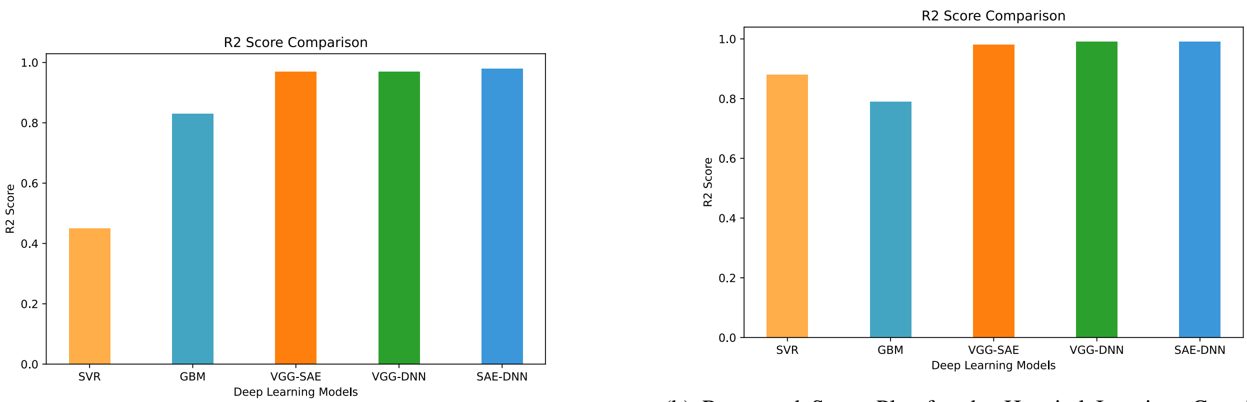
FIGURE 7. **MAE Comparison of ML techniques with hybrid models on healthcare providers and hospital inpatient cost transparency datasets.**



(a) R-squared Score Plot for the Healthcare Providers Dataset

(b) R-squared Score Plot for the Hospital Inpatient Cost Transparency Dataset

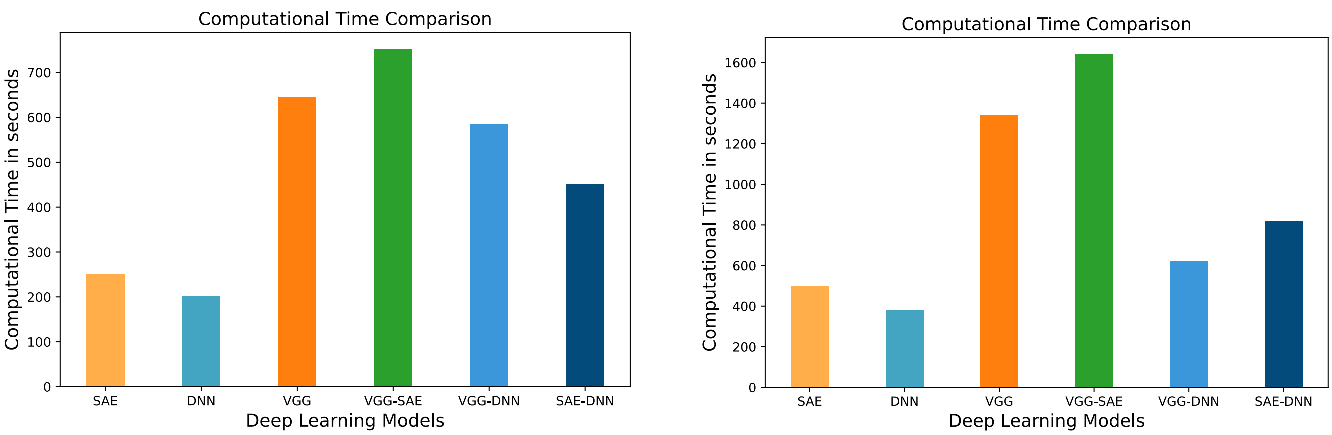
FIGURE 8. R-squared Score Comparison of Individual and Hybrid DL Models for the Healthcare Providers and Hospital Inpatient Cost Transparency Datasets.



(a) R-squared Score Plot for the Healthcare Providers Dataset

(b) R-squared Score Plot for the Hospital Inpatient Cost Transparency Dataset

FIGURE 9. R-squared Comparison of ML techniques with hybrid models on healthcare providers and hospital inpatient cost transparency datasets.

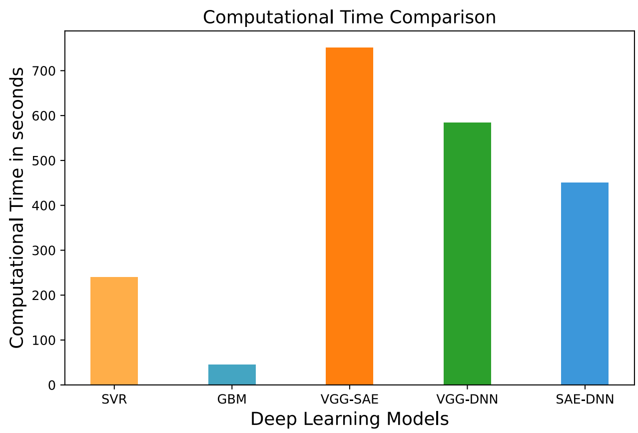


(a) Time Plot for the Healthcare Providers Dataset

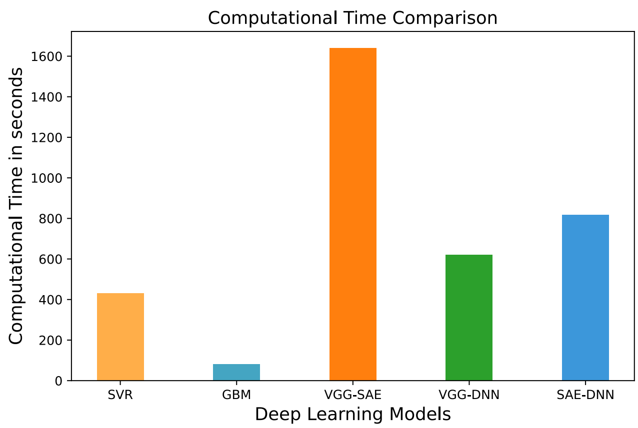
(b) Time Plot for the Hospital Inpatient Cost Transparency Dataset

FIGURE 10. Time comparison of individual and hybrid DL models for the healthcare providers and hospital inpatient cost transparency datasets.

are presented. According to the summary plot, the first two features, Average Medicare Payment Amount and Average Medicare Allowed Amount, show the maximum contribution on the predictions of the model. The remaining features have little effect on the outcomes of the model.

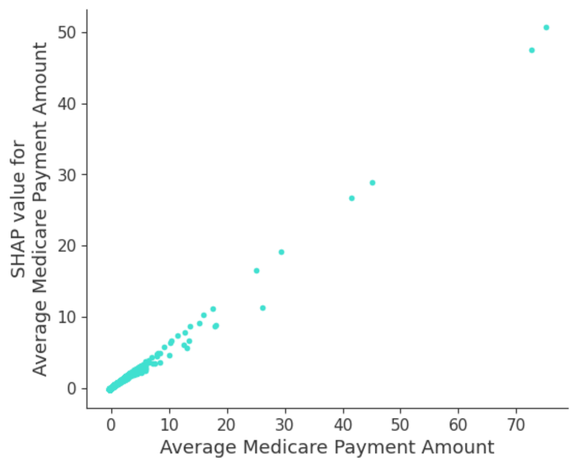


(a) Time Plot for the Healthcare Providers Dataset



(b) Time Plot for the Hospital Inpatient Cost Transparency Dataset

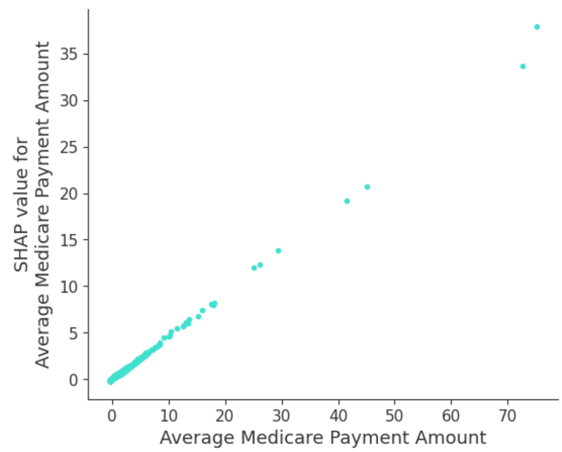
FIGURE 11. **Time comparison of ML techniques with hybrid models on healthcare providers and hospital inpatient cost transparency datasets.**



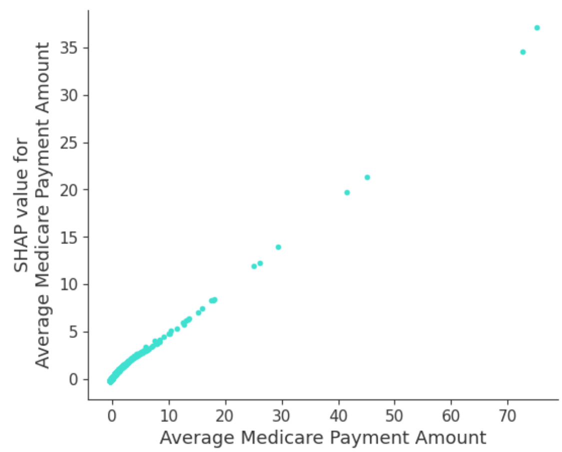
(a) Dependence Plot of VGG-DNN Model

(b) Dependence Plot of VGG-DNN Model

FIGURE 12. **Dependency plots of VGG-DNN model for average medicare payment amount feature.**



(a) Dependence Plot of VGG-SAE Model



(b) Dependence Plot of SAE-DNN Model

FIGURE 13. **Dependency plots of VGG-SAE, and SAE-DNN models for average medicare payment amount feature.**

The Number of Distinct Medicare Beneficiary/Per Day Services feature shows a negative effect on the predictions. In the bar plot, the first two features exhibit maximum impact on the prediction of VGG-SAE. In the force plot, the feature in the red color such as Average Submitted Charge Amount has a negative effect on the prediction. The values in the blue color such as Average Medicare Payment Amount and Average Medicare Allowed Amount exhibit a positive effect

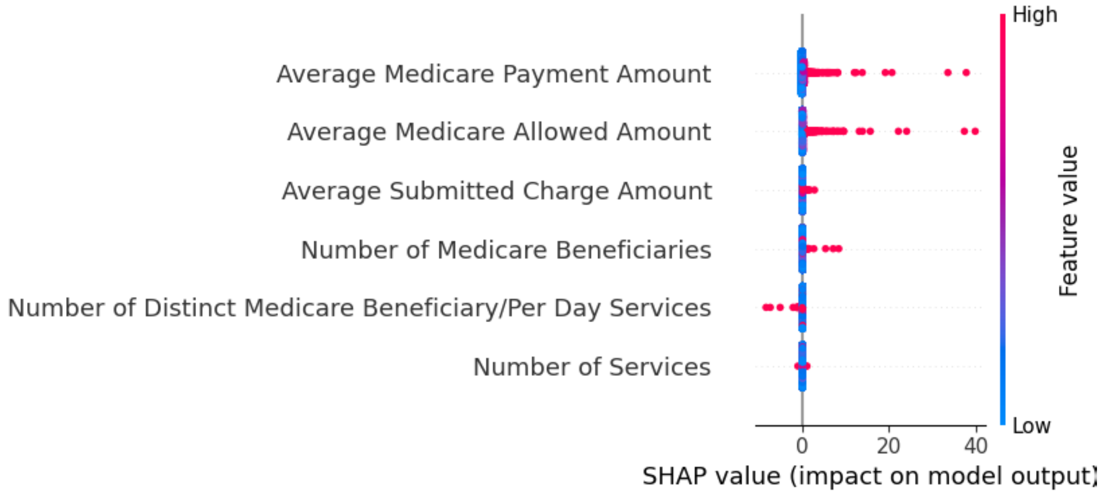


FIGURE 14. Summary plot of VGG-SAE model.

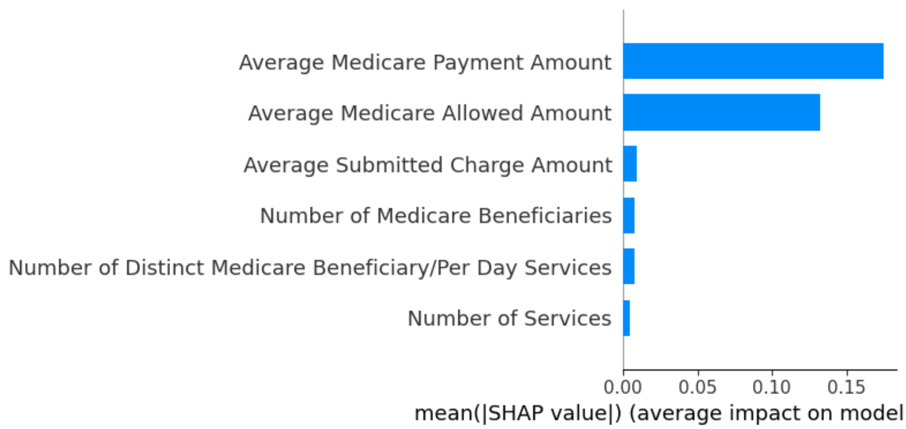


FIGURE 15. Bar plot of VGG-SAE model.

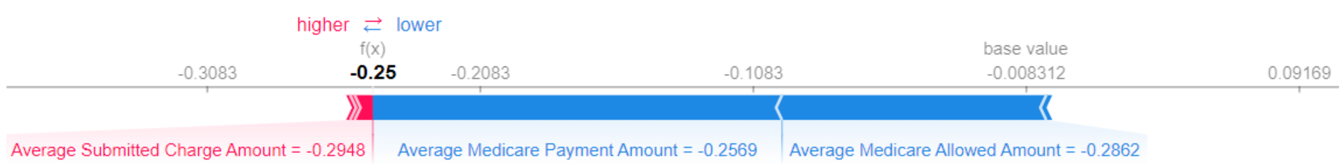


FIGURE 16. Force plot of VGG-SAE model.

on the prediction of VGG-SAE. In Figures 17, 18 and 19, the plot results of VGG-DNN are mostly similar to the VGG-SAE model. The only difference in the summary plot is that the SHAP value of the feature Average Medicare Allowed Amount is less than the SHAP values in VGG-SAE. In Figures 20, 21 and 22, the SHAP values of SAE-DNN prioritize the first two features such as Average Medicare Payment Amount and Average Medicare Allowed Amount. while, other than the first two features, the impact of other features on the prediction is negligible. From the force plot of SAE-DNN, it is visible that the Average Submitted Charge Amount has a negative impact on prediction. While the Average Medicare Payment Amount and Average Medicare Allowed Amount have a positive impact on the predictions.

VI. CONCLUSION

In this paper, new hybrid models such as SAE-DNN, VGG-DNN, and VGG-SAE are proposed to predict the cost estimation of healthcare providers. In these hybrid combinations, deep models are specifically used to reduce the dimension-

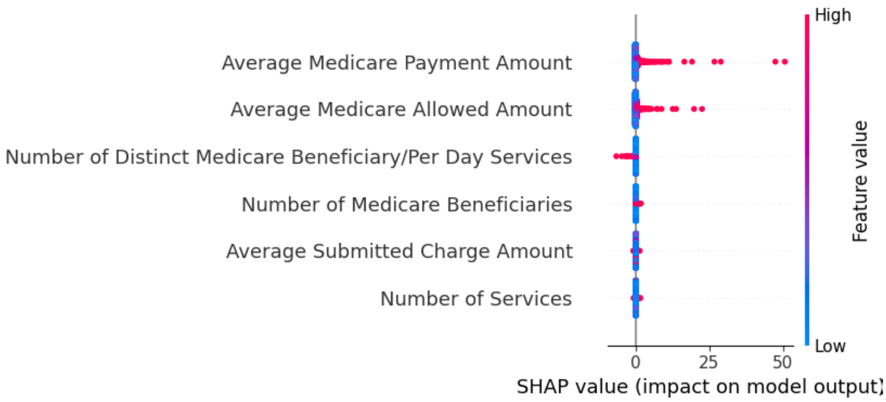


FIGURE 17. Summary plot of VGG-DNN model.

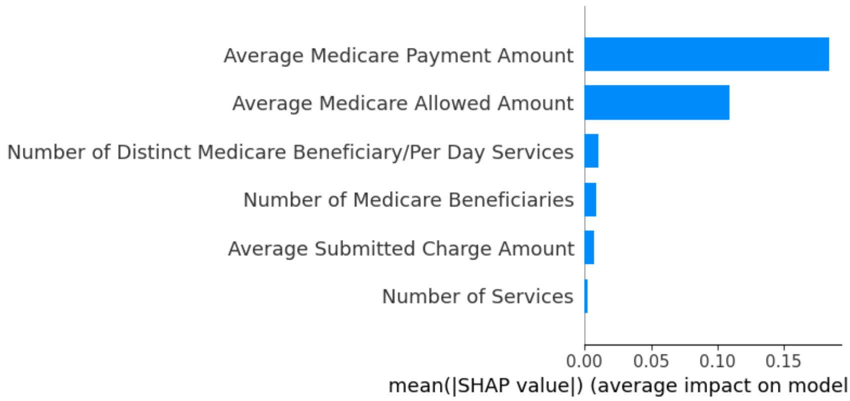


FIGURE 18. Bar plot of VGG-DNN model.

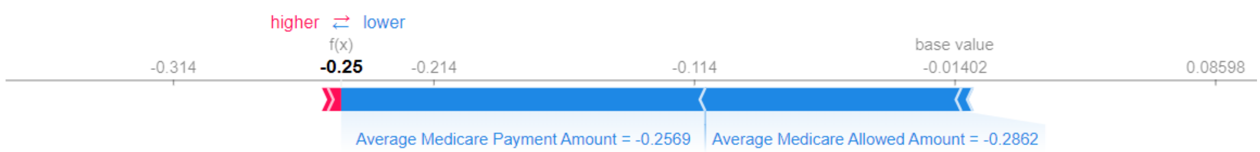


FIGURE 19. Force plot of VGG-DNN model.

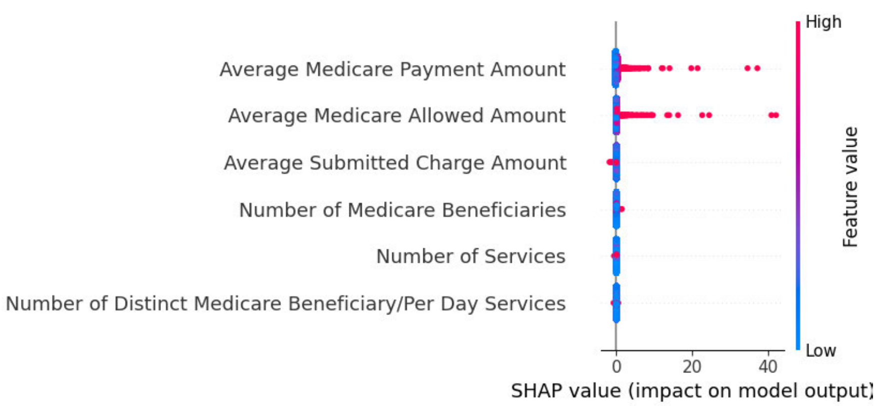


FIGURE 20. Summary plot of SAE-DNN model.

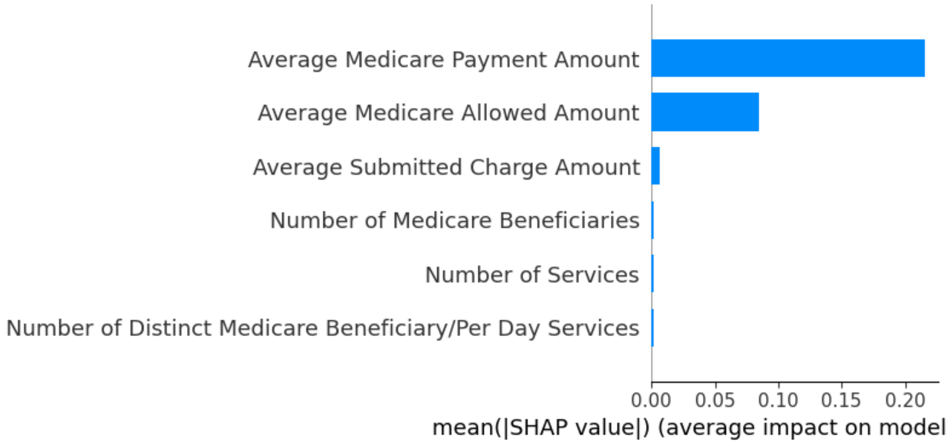


FIGURE 21. Bar plot of SAE-DNN model.

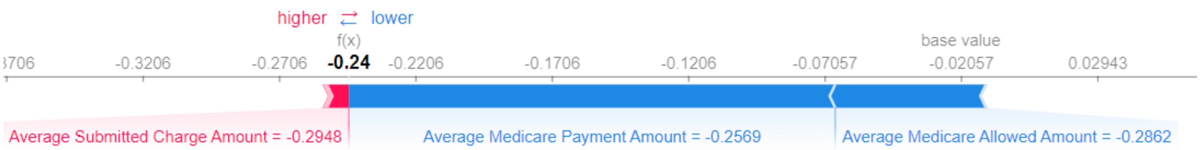


FIGURE 22. Force plot of SAE-DNN model.

ality reduction of the training dataset and extract the feature vectors using the neural network architecture. It contributes to performance by selecting only relevant features of input features. After the dimensionality reduction, the prediction of the cost estimation of the healthcare providers is performed. Random search is utilized to obtain the optimal hyper-parameters of each hybrid model. To check the robustness of our proposed approach, the proposed models are validated on two different datasets. SHAPley is used to explain the importance of each feature in the prediction performed by the hybrid models. The three combinations of hybrid models help the patients in healthcare resource allocation and resource management. SAE-DNN, VGG-DNN, and VGG-SAE are compared with other DL models such as SAE, DNN, and VGG. With the hyper-parameter tuning of the Healthcare Providers Dataset, VGG-SAE achieved MSE of 0.01, RMSE of 0.13, MAE of 0.02, and R-squared of 0.98. VGG-DNN achieved MSE of 0.01, RMSE of 0.12, MAE of 0.02, and R-squared of 0.99. SAE-DNN achieved MSE of 0.01, RMSE of 0.11, MAE of 0.02, and R-squared of 0.99. With the hyper-parameter tuning of the Hospital Inpatient Cost Transparency Dataset, VGG-SAE achieved MSE of 0.007, RMSE of 0.08, MAE of 0.03, R-squared of 0.99, and execution time of 1680 seconds. VGG-DNN achieved MSE of 0.0006, RMSE of 0.08, MAE of 0.03, R-squared of 0.99, and execution time of 645 seconds. SAE-DNN achieved MSE of 0.003, RMSE of 0.06, MAE of 0.02, R-squared of 0.99, and execution time of 850 seconds. SAE-DNN, VGG-DNN, and VGG-SAE outperformed individual DL models concerning different evaluation parameters such as MSE, RMSE, MAE and R-squared score.

VII. FUTURE WORK

The DL models perform more accurately in healthcare than ML techniques. VGG-SAE, VGG-DNN, and SAE-DNN predict the cost of healthcare providers more accurately than individual models and ML techniques such as VGG, SAE, DNN, SVR, and GBR. However, these techniques have some limitations that will be handled in the future in a better way. SAE lacks to learn the information as much as possible regardless of relevant data. Due to multiple AE layers, it takes more time to process the dataset. SAE is very sensitive to a noisy or unprocessed dataset that results in wrongly predicting the cost of healthcare providers. The latent space in the bottleneck is very narrow, which restricts further training of the model. During the encoding and decoding process, SAE loses the important information that may be important for the prediction. DNN faces the vanishing gradient problem that takes more time to train the model. Also, due to the addition of multiple hidden layers in the DNN, it takes more computational time and resources. VGG takes more time to train the model due to the massive addition of layers in the model. VGG needs to train 138 million parameters that lead to the vanishing and exploding gradient problems. SVR performs wrong prediction on large datasets. It is more sensitive to noisy values of the dataset that affect the prediction. Also, it underfits when the feature values exceed the training data points. GBR lacks the ability to consume more computational resources due to a massive number of decision trees. In addition, GBR does not perform better on large datasets and may overfit when the model is complex.

ACKNOWLEDGMENT

The authors extend their appreciation to the Researchers Supporting Program for Project number (RSPD2023R648), at King Saud University for supporting this research project.

REFERENCES

[1] R. L. Sanz and P. Leon-Sanz, "Modeling health data using machine learning techniques applied to financial management predictions," *Appl. Sci.*, vol. 12, no. 23, p. 12148, Nov. 2022.

[2] J. M. Hyer, A. Z. Paredes, S. White, A. Ejaz, and T. M. Pawlik, "Assessment of utilization efficiency using machine learning techniques: A study of heterogeneity in preoperative healthcare utilization among super-utilizers," *Amer. J. Surgery*, vol. 220, no. 3, pp. 714–720, Sep. 2020.

[3] L. Luo, X. Yu, Z. Yong, C. Li, and Y. Gu, "Design comorbidity portfolios to improve treatment cost prediction of asthma using machine learning," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2237–2247, Jun. 2021.

[4] J. S. Berger, L. Haskell, W. Ting, F. Lurie, S.-C. Chang, L. A. Mueller, K. Elder, K. Rich, C. Crivera, J. R. Schein, and V. Alas, "Evaluation of machine learning methodology for the prediction of healthcare resource utilization and healthcare costs in patients with critical limb ischemia—Is preventive and personalized approach on the horizon?" *EPMA J.*, vol. 11, no. 1, pp. 53–64, Mar. 2020.

[5] M. Z. Younis, S. Jaber, A. R. Mawson, and M. Hartmann, "Estimating the unit costs of public hospitals and primary healthcare centers," *Int. J. Health Planning Manage.*, vol. 28, no. 4, pp. 320–332, Oct. 2013.

[6] D. Anderson, M. V. Bjarnadottir, and Z. Nenova, "Machine learning in healthcare: Operational and financial impact," in *Innovative Technology at the Interface of Finance and Operations*, vol. 1. Heidelberg, Germany: Springer, 2022, pp. 153–174.

[7] R. Khalid, N. Javaid, F. A. Al-Zahrani, K. Aurangzeb, E. U. H. Qazi, and T. Ashfaq, "Electricity load and price forecasting using Jaya-long short term memory (JLSTM) in smart grids," *Entropy*, vol. 22, no. 1, p. 10, 2019.

[8] M. Adil, N. Javaid, U. Qasim, I. Ullah, M. Shafiq, and J.-G. Choi, "LSTM and bat-based RUSBoost approach for electricity theft detection," *Appl. Sci.*, vol. 10, no. 12, p. 4378, Jun. 2020.

[9] B. Kabir, U. Qasim, N. Javaid, A. Aldegheishem, N. Alrajeh, and E. A. Mohammed, "Detecting nontechnical losses in smart meters using a MLP-GRU deep model and augmenting data via theft attacks," *Sustainability*, vol. 14, no. 22, p. 15001, Nov. 2022.

[10] J. M. Karnuta, J. L. Churchill, H. S. Haeberle, B. U. Nwachukwu, S. A. Taylor, E. T. Ricchetti, and P. N. Ramkumar, "The value of artificial neural networks for predicting length of stay, discharge disposition, and inpatient costs after anatomic and reverse shoulder arthroplasty," *J. Shoulder Elbow Surgery*, vol. 29, no. 11, pp. 2385–2394, Nov. 2020.

[11] S. A. Bini, "Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care?" *J. Arthroplasty*, vol. 33, no. 8, pp. 2358–2361, Aug. 2018.

[12] T. G. Myers, P. N. Ramkumar, B. F. Ricciardi, K. L. Urish, J. Kipper, and C. Ketonis, "Artificial intelligence and orthopaedics: An introduction for clinicians," *J. Bone Joint Surg. Amer.*, vol. 102, no. 9, pp. 830–840, 2020.

[13] P. N. Ramkumar, J. M. Karnuta, S. M. Navarro, H. S. Haeberle, R. Iorio, M. A. Mont, B. M. Patterson, and V. E. Krebs, "Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: Development and validation of a deep learning model," *J. Arthroplasty*, vol. 34, no. 10, pp. 2228–2234.e1, Oct. 2019.

[14] Pamir, N. Javaid, M. Akbar, A. Aldegheishem, N. Alrajeh, and E. A. Mohammed, "Employing a machine learning boosting classifiers based stacking ensemble model for detecting non technical losses in smart grids," *IEEE Access*, vol. 10, pp. 121886–121899, 2022.

[15] A. Ullah, N. Javaid, M. Asif, M. U. Javed, and A. S. Yahaya, "AlexNet, AdaBoost and artificial bee colony based hybrid model for electricity theft detection in smart grids," *IEEE Access*, vol. 10, pp. 18681–18694, 2022.

[16] R. Bhardwaj, A. R. Nambiar, and D. Dutta, "A study of machine learning in healthcare," in *Proc. IEEE 41st Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, vol. 2, Jul. 2017, pp. 236–241.

[17] D. K. K. Goyal and A. H. Paray, "A survey of different approaches of machine learning in healthcare management system," *Int. J. Adv. Netw. Appl.*, vol. 11, no. 3, pp. 4270–4276, 2019.

[18] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Literature review: Machine learning techniques applied to financial market prediction," *Exp. Syst. Appl.*, vol. 124, pp. 226–251, Jun. 2019.

[19] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Inf. Fusion*, vol. 66, pp. 170–183, Feb. 2021.

[20] K. Jiang, J. Tang, Y. Wang, C. Qiu, Y. Zhang, and C. Lin, "EEG feature selection via stacked deep embedded regression with joint sparsity," *Frontiers Neurosci.*, vol. 14, p. 829, Aug. 2020.

[21] F. Wang, L. P. Casalino, and D. Khullar, "Deep learning in medicine-promise, progress, and challenges," *JAMA Internal Med.*, vol. 179, no. 3, pp. 293–294, 2019.

[22] K. Teo, C. W. Yong, F. Muhamad, H. Mohafez, K. Hasikin, K. Xia, P. Qian, S. Dhanalakshmi, N. P. Utama, and K. W. Lai, "The promise for reducing healthcare cost with predictive model: An analysis with quantized evaluation metric on readmission," *J. Healthcare Eng.*, vol. 2021, pp. 1–10, Nov. 2021.

[23] R. Norat, A. S. Wu, and X. Liu, "Genetic algorithms with self-adaptation for predictive classification of medicare standardized payments for physical therapists," *Exp. Syst. Appl.*, vol. 218, May 2023, Art. no. 119529.

[24] M. A. Morid, O. R. L. Sheng, K. Kawamoto, and S. Abdelrahman, "Learning hidden patterns from patient multivariate time series data using convolutional neural networks: A case study of healthcare cost prediction," *J. Biomed. Informat.*, vol. 111, Nov. 2020, Art. no. 103565.

[25] X. Zeng, S. Lin, and C. Liu, "Multi-view deep learning framework for predicting patient expenditure in healthcare," *IEEE Open J. Comput. Soc.*, vol. 2, pp. 62–71, 2021.

[26] S. Kaushik, A. Choudhury, S. Natarajan, L. A. Pickett, and V. Dutt, "Medicine expenditure prediction via a variance-based generative adversarial network," *IEEE Access*, vol. 8, pp. 110947–110958, 2020.

[27] K. Yu, Z. Yang, C. Wu, Y. Huang, and X. Xie, "In-hospital resource utilization prediction from electronic medical records with deep learning," *Knowl.-Based Syst.*, vol. 223, Jul. 2021, Art. no. 107052.

[28] A. Abbas, J. Mosseri, J. R. Lex, J. Toor, B. Ravi, E. B. Khalil, and C. Whyne, "Machine learning using preoperative patient factors can predict duration of surgery and length of stay for total knee arthroplasty," *Int. J. Med. Informat.*, vol. 158, Feb. 2022, Art. no. 104670.

[29] C. Puri, G. Kooijman, B. Vanrumste, and S. Luca, "Forecasting time series in healthcare with Gaussian processes and dynamic time warping based subset selection," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 12, pp. 6126–6137, Dec. 2022.

[30] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020.

[31] F. Desai, D. Chowdhury, R. Kaur, M. Peeters, R. C. Arya, G. S. Wander, S. S. Gill, and R. Buyya, "HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing," *Internet Things*, vol. 17, Mar. 2022, Art. no. 100485.

[32] C. Gu, C. Dai, X. Shi, Z. Wu, and C. Chen, "A cloud-based deep learning model in heterogeneous data integration system for lung cancer detection in medical industry 4.0," *J. Ind. Inf. Integr.*, vol. 30, Nov. 2022, Art. no. 100386.

[33] K. N. Qureshi, S. Din, G. Jeon, and F. Piccialli, "An accurate and dynamic predictive model for a smart M-health system using machine learning," *Inf. Sci.*, vol. 538, pp. 486–502, Oct. 2020.

[34] S. Cheon, J. Kim, and J. Lim, "The use of deep learning to predict stroke patient mortality," *Int. J. Environ. Res. Public Health*, vol. 16, no. 11, p. 1876, May 2019.

[35] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, "Heart failure: Diagnosis, severity estimation and prediction of adverse events through machine learning techniques," *Comput. Structural Biotechnol. J.*, vol. 15, pp. 26–47, 2017.

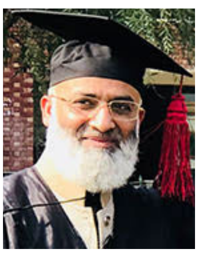
[36] H. K. Bhuyan, V. Ravi, B. Brahma, and N. K. Kamila, "Disease analysis using machine learning approaches in healthcare system," *Health Technol.*, vol. 12, no. 5, pp. 987–1005, Sep. 2022.

[37] *Healthcare Providers Data For Anomaly Detection*. Accessed: Aug. 8, 2023. [Online]. Available: https://www.kaggle.com/datasets/tamilsel/healthcare-providers-data

[38] *Health Data NY—New York State*. Accessed: Aug. 8, 2023. [Online]. Available: https://health.data.ny.gov/w/7dtz-qxmr/fbc6-cypp?cur=rdB1zBiIz70&from=zmwCSaVUPwa

[39] M. Yu, T. Quan, Q. Peng, X. Yu, and L. Liu, "A model-based collaborate filtering algorithm based on stacked AutoEncoder," *Neural Comput. Appl.*, vol. 34, no. 4, pp. 2503–2511, Feb. 2022.
[40] X. Yuan, C. Ou, Y. Wang, C. Yang, and W. Gui, "Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE," *Neurocomputing*, vol. 396, pp. 375–382, Jul. 2020.
[41] Accessed: May 2, 2023. [Online]. Available: https://medium.com/@venkatakrishna.jonnalagadda/sparse-stacked-and-variational-autoencoder-efe5bfe73b64#:~:text=A%20stacked%20autoencoder%20is%20a,of%20the%20successive%20hidden%20layer
[42] J. Du and Y. Xu, "Hierarchical deep neural network for multivariate regression," *Pattern Recognit.*, vol. 63, pp. 149–157, Jul. 2017.
[43] D. Ou, K. Tan, J. Lai, X. Jia, X. Wang, Y. Chen, and J. Li, "Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction," *Geoderma*, vol. 385, Mar. 2021, Art. no. 114875.
[44] Y. Zhang, Y. Xie, Y. Zhang, J. Qiu, and S. Wu, "The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity," *Bull. Eng. Geol. Environ.*, vol. 80, pp. 5053–5060, Apr. 2021.
[45] *Deep Probabilistic Modeling (III). Artificial Neural Networks*. Accessed: May 8, 2023. [Online]. Available: https://andresmasegosa.github.io/DeepProb-dnn/
[46] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 33, pp. 4741–4753, Nov. 2021.
[47] Accessed: May 2, 2023. [Online]. Available: https://www.baeldung.com/cs/cnn-feature-map
[48] X. Li, K. Zhang, W. Li, Y. Feng, and R. Liu, "A two-stage transfer regression convolutional neural network for bearing remaining useful life prediction," *Machines*, vol. 10, no. 5, p. 369, 2022.
[49] *Fully 2D Connected Layers VGG 2D, Each Corresponds to One Class*. Accessed: May 5, 2023. [Online]. Available: https://iq.opengenus.org/vgg-11/#:~:text=Fully%2DConnected%20Layers%20%3A%20VGG%2D,each%20corresponds%20to%20one%20class
[50] D. Parbat and M. Chakraborty, "A Python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons Fractals*, vol. 138, Sep. 2020, Art. no. 109942.
[51] X. Chen, L. Huang, D. Xie, and Q. Zhao, "EGBMMDA: Extreme gradient boosting machine for MiRNA-disease association prediction," *Cell Death Disease*, vol. 9, no. 1, p. 3, 2018.
[52] C. A. Huber, S. Schneeweiss, A. Signorell, and O. Reich, "Improved prediction of medical expenditures and health care utilization using an updated chronic disease score and claims data," *J. Clin. Epidemiol.*, vol. 66, no. 10, pp. 1118–1127, 2013.
[53] *What is a Good RMSE Value? Simply Explained*. Accessed: May 7, 2023. [Online]. Available: https://stephenallwright.com/good-rmse-value/

MUHAMMAD HAMZA RAFIQUE BHATTI received the bachelor's degree in software engineering from COMSATS University Islamabad, Islamabad, in 2021, and the master's degree in software engineering, with a specialization in deep learning and healthcare sector using AI from the Communications Over Sensors (ComSens) Laboratory, Department of Computer Science, COMSATS University Islamabad, in 2023, under the supervision of Dr. Nadeem Javaid. He is currently a Research Associate with the ComSens Laboratory. His research interests include artificial intelligence, machine learning, deep learning, and healthcare.



NADEEM JAVAID (Senior Member, IEEE) received the bachelor's degree in computer science and physics from Gomal University, Dera Ismail Khan, Pakistan, in 1995, the master's degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 1999, and the Ph.D. degree from the University of Paris-Est, France, in 2010. He has 25 years of teaching and research experience. He was a Visiting Professor with the University of Technology Sydney, Australia. He is currently a tenured Professor and the Founding Director of the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad Campus. He has supervised 187 master's and 30 Ph.D. theses. He has authored over 950 papers in technical journals and international conferences. His research interests include energy optimization in smart/microgrids and wireless sensor networks, using data analytics and blockchain. He was a recipient of the Best University Teacher Award (BUTA'16) from the Higher Education Commission (HEC) of Pakistan, in 2016; and the Research Productivity Award (RPA'17) from the Pakistan Council for Science and Technology (PCST), in 2017. He is an Editor of *Sustainable Cities and Society*.

BABAR MANSOOR received the B.Sc. degree in computer engineering from the COMSATS Institute of Information Technology (CIIT), Abbottabad, Pakistan, in 2005, the M.Sc. degree in microelectronic systems from the University of Liverpool, U.K., in 2007, and the Ph.D. degree in electrical engineering from COMSATS University Islamabad (CUI), Islamabad, Pakistan, in 2018. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, CUI. His research interests include wireless communications, signal processing, and MIMO communication systems.

NABIL ALRAJEH received the Ph.D. degree in biomedical informatics engineering from Vanderbilt University, Nashville, TN, USA. He was a Senior Advisor with the Ministry of Higher Education, where he was implementing development programs, including educational affairs, strategic planning, and research and innovation. He was a member of the boards of trustees for five private universities in Saudi Arabia. He is currently a Professor of health informatics with King Saud University. His research interests include e-health applications, hospital information systems, telemedicine, healthcare applications of smart cities, and wireless sensor networks.

MUHAMMAD ASLAM received the M.S. degree from the Department of Electrical Engineering (Networking), COMSATS University Islamabad, Islamabad, Pakistan, and the Ph.D. degree from the School of Computer Science and Technology, Dalian University of Technology, Dalian, China, in June 2018, under the supervision of Prof. Xiaopeng Hu. He is currently a Lecturer with the University of the West of Scotland. Before that, he was a Postdoctoral Research Scholar with the School of Cyber Science and Engineering, Wuhan University, under the International Postdoctoral Exchange Fellowship Program. His research interests include machine learning, cyber security, software-defined networking, wireless sensor networks, and networks security.

MUHAMMAD ASAD (Member, IEEE) received the B.S. degree in telecommunication and networking (BSTN) from COMSATS University Islamabad, Wah Campus, Pakistan, June 2014, the M.S. degree in computer science (MSCS) from the Dalian University of Technology, Dalian, China, in June 2018, and the Ph.D. degree in computer science from the Department of Computer Science, Nagoya Institute of Technology, Japan, in June 2022. He is currently a Project Researcher with the Graduate School of Information Science and Technology, The University of Tokyo, Japan. He received the Chinese Government Scholarship (CSC), from 2015 to 2018. He also received the Japanese Government Scholarship (MEXT), from 2019 to 2022. His primary research interests include federated learning, deep learning, software-defined networking (SDN), wireless sensor networks (WSNs), vehicular ad-hoc networks (VANets), and the Internet of Things (IoT).