A Deep Analysis for Medical Emergency Missing Value Imputation

Md Faisal Kabir[a] and Sven Tomforde[b]

Institute of Computer Science, Department of Intelligent Systems, University of Kiel, Christian-Albrechts-Platz 4, Kiel, Germany

Keywords: Medical Emergency System, Missing Value, Data Imputation, Autoencoder, Deep Learning.

Abstract: The prevalence of missing data is a pervasive issue in the medical domain, necessitating the frequent deployment of various imputation techniques. Within the realm of emergency medical care, multiple challenges have been addressed, and solutions have been explored. Notably, the development of an AI assistant for telenotary service (TNA) encounters a significantly higher frequency of missing values compared to other medical applications, with these values missing completely at random. In response to this, we compare several traditional machine learning algorithms with denoising autoencoder and denoising LSTM autoencoder strategies for imputing numerical (continuous) missing values. Our study employs a genuine medical emergency dataset, which is not publicly accessible. This dataset exhibits a significant class imbalance and includes numerous outliers representing rare occurrences. Our findings indicate that the denoising LSTM autoencoder outperforms the conventional approach.

1 INTRODUCTION

Medical emergencies represent a paramount concern, posing substantial challenges not only in addressing patients' life-threatening conditions but also in the prompt assessment and treatment initiation faced by healthcare professionals. This process is inherently time-intensive. Therefore, an existing telenotary service (TeleNotarzt, TNA) (Aachen, 2023) is being undertaken to enhance support during medical emergencies.

The successful categorization of diseases in future emergency medical applications requires a substantial dataset. Our dataset, sourced from a decade of TNA operations in Aachen, Germany, includes measurement data records for patients and their emergency cases. However, numerous patient samples lack essential parameters, potentially leading to inaccurate classification results. Furthermore, the dataset displays a notable class imbalance and incorporates outliers reflective of genuine medical events, which are crucial for understanding specific diseases. This study aims to investigate missing value imputation techniques that effectively handle minority class instances while accurately identifying outliers linked to rare disease events.

Within this research paper, we propose two denoising autoencoder models, namely the conventional denoising autoencoder and the denoising LSTM autoencoder, with the primary objective of addressing missing value imputation. Subsequently, we perform a comparative analysis of both models with an emphasis on optimization. Additionally, our proposed model is systematically evaluated against various machine learning models and statistical techniques, including mean imputations, K-nearest neighbors (KNN), Iterative Imputer (I.Imp), Decision Tree (D.Tree), and Linear Regression (LR). This is the first study to complete the missing imputation for the emergency medical dataset, a corresponding real-world-based. With this paper, we aim to answer the question of which techniques are most suitable for this kind of scenario and which challenges need to be tackled.

The paper is organized as follows: Section 2 reviews existing literature on missing data patterns, techniques, and methodologies. Section 3 discusses current state-of-the-art developments. Section 4 explains the formulation of proposed models, and Section 5 details data description and preparation. Model evaluation is covered in Section 6, while Section 7 analyzes experimental outcomes, presenting insights and findings. Finally, Section 8 summarises the paper and outlines future research directions.

[a] https://orcid.org/0009-0000-2354-2193
[b] https://orcid.org/0000-0002-5825-8915

Kabir, M. and Tomforde, S.
A Deep Analysis for Medical Emergency Missing Value Imputation.
DOI: 10.5220/0012457300003636
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 3*, pages 1229-1236
ISBN: 978-989-758-680-4; ISSN: 2184-433X
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

2 BACKGROUND

Missing values are a common problem in a dataset. The causes of missing values are the data entry mechanism not working, updating the new system, human error, system error, faulty measurements, etc. In this case, the critical issue needs values for future product improvement. So, it is necessary to know 1) types of missing value, 2) strategies of missing value imputation and 3) imputation techniques.

2.1 Types of Missing Values

In situations where data are classified as Missing Completely at Random (MCAR), there is an absence of any discernible connection between missing and non-missing values. In other words, missing and non-missing variables contain no relationship. The missing value placed in a dataset is entirely random (Emmanuel et al., 2021; Hameed and Ali, 2023). So, the probability is represented as: $P(X_j|X_j,X_i) = P(X_j)$. This equation indicates that the probability of missingness (X_j) is not influenced by the observed data (X_i) or the missing data (X_j), as the conditional probability is equal to the unconditional probability (Emmanuel et al., 2021; Hameed and Ali, 2023).

Missing at Random (MAR) considers if the missing data is connected to observed variables and has a relationship between the subset of that variable. For example, if we consider X_i as an observed variable, which is the person's name, X_i is age, X_m is monthly income, and other variables are different parameters. Here X_j is missing variables (Emmanuel et al., 2021; Hameed and Ali, 2023). We can write that the probability of missing is $P(X_m|X_m,X_i) = P(X_m|X_i)$.

When the missing variable depends on itself and the observed variable, we call it Missing Not at Random (MNAR) (Emmanuel et al., 2021; Hameed and Ali, 2023). This can be expressed as: $P(X_m|X_m,X_i) = f(X_mX_i)$.

2.2 Missing Data Handling

AI missing imputation divides into two parts: machine learning and deep learning. Deep learning is a subsection of machine learning, but deep learning is one of the best approaches to complete missing values. The Deep learning approach mainly focuses on model-based and predicts missing values to achieve the target dataset. Deep learning uses a neural network with multiple layers to discover the dataset's complex pattern. Conversely, deep learning is complicated to construct a model and requires high computational capacity (Emmanuel et al., 2021; Liu et al., 2023a; Sun et al., 2023; Liu et al., 2023b).

2.3 Missing Imputation Techniques

Machine learning missing value imputation techniques based on Classification and clustering models. Bayesian, neural network, decision tree, random forest and support vector machine are considered classification-based models. On the other hand, k-means, subspace, self-organising map, and kernel are clustering-based models. To complete missing imputation, autoencoder (AE), variational autoencoder (VAE), generative adversarial networks (GANs), Long short-term memory (LSTM) networks, and transformers are used in deep learning (Adhikari et al., 2022; Emmanuel et al., 2021; Liu et al., 2023a; Sun et al., 2023).

3 RELATED WORK

Recently, many methods have been employed for missing data imputations across diverse data types spanning various sectors. This section mainly concentrates on the application of these methods in the context of medical and healthcare datasets. The current state of the art reveals the utilization of multiple models based on deep learning techniques for addressing missing data across various data types.

In the medical domain, missing imputation techniques are extensively utilized for tabular static data, signals, temporal data, images, and genetic and genomic data. Deep learning models have demonstrated superior accuracy compared to alternative algorithms. Typically, missing data imputation serves as a critical step in data preprocessing, with the ultimate objective often being prediction or classification analysis. Concurrently, various statistical and machine learning models are also employed to assess imputation accuracy for subsequent processes (Liu et al., 2023a).

Notably, multilayer perceptron (MLP) with gradient descent emerges as a successful method for imputing missing values, showcasing superior performance compared to other statistical methods such as mode, random, hot-deck, KNN, decision tree, and random forest (Pan et al., 2022). GapNet (Chang et al., 2022), another approach rooted in the MLP model, exhibits similar imputation performance as MLP with gradient descent. MLP models, known for their effectiveness in handling static data, also demonstrate noteworthy performance in genetic datasets, exemplified by deepMC (Mongia et al., 2020).

For temporal data, models such as Long Short-Term Memory (LSTM), Recurrent Neural Network

(RNN), and Gated Recurrent Unit (GRU) are specifically designed and proficient in handling time series missing values. To predict hypertensive disorder in pregnancy (HDP) using pregnancy examination data, a Bidirectional LSTM model outperforms cubic spline interpolation, KNN filling, the LSTM model, and the ST-MVL model (Lu et al., 2023). On the other hand, a forward-to-backward bidirectional model employing missing imputation demonstrates enhanced accuracy in predicting Alzheimer's disease compared to other methods (Ho et al., 2022).

Within the medical domain, autoencoders are extensively employed for missing value imputation. An illustrative experiment utilized a sample autoencoder to impute missing values using a dataset with randomly masked values in the missing positions. This approach was compared with mean imputation, revealing superior results in favour of the autoencoder method over mean imputation (Macias et al., 2021). Additionally, employing a random mask as an auxiliary input proved beneficial in assisting the network to discern between missing and zero features. In this context, autoencoders demonstrated superior performance across five distinct datasets, namely Breast Cancer Wisconsin (WDBC), Parkinson's disease, diabetic data, Indian Liver Patient Dataset (ILPD), and National Surgical Quality Improvement Program (NSQIP). This performance surpassed that of the support vector machine (SVM), k-nearest neighbour (KNN), artificial neural network (ANN), linear regression (LR), and random forest (RF) (Kabir and Farrokhvar, 2022).

A stacked denoising autoencoder (DAE) was implemented with a low-dimensional representation. The weight matrix of the last layer in the encoder was utilized as the transpose weight matrix in the first layer of the decoder (Abiri et al., 2019). In the Framingham Heart dataset, the process of missing imputation was conducted using a DAE as the generator within an Improved Generative Adversarial Network (IGAN). Before inputting the data into the model, missing values were filled using K-Nearest Neighbors (KNN). Notably, IGAN demonstrated superior performance when compared to Simple imputation, KNN, MissForest, Neighborhood Aware Autoencoder (NAA), and Improved Neighborhood Aware Autoencoder (INAA) (Psychogyios et al., 2022).

The current state of the art reveals numerous studies addressing missing imputation in various medical datasets. However, no existing work has specifically focused on addressing the challenge of emergency medical missing data imputation. In response, we study two autoencoder models—the Denoising Autoencoder and the Denoising LSTM Autoencoder for the imputation of emergency medical data.

4 PROPOSED MISSING VALUE IMPUTATION

4.1 Autoencoder

An autoencoder (AE) comprises an encoder and a decoder. The encoder processes the input data and generates an output that represents a reduced-dimensional vector space, commonly referred to as the latent space or information bottleneck. The mathematical functions governing the encoder and decoder are denoted as $f(x) = z$ and $g(z) = x'$. Let us provide a more detailed elaboration of these functions. In the context of model construction for the encoder, $w1$ and $b1$ represent the weight matrix and bias of the first layer, respectively. Here, x and f denote the input and activation functions, leading to the presentation of the output as z. (Wang et al., 2016; Chen and Guo, 2023) The equation can be expressed as follows: $z = f(w \times x + b)$. Same as for the decoder, w', b',g and x' are weight, bias, activation function of decoder output layer and reconstruction data presented. z is the input of the decoder. So the equation presented as: $x' = g(w' \times z + b')$. Training an autoencoder loss function is also an important issue. The loss function calculates how input data is reconstructed. To define the loss function, the following expression is used: $L(x, g(f(x)))$

4.2 Denoising Autoencoder

The denoising autoencoder (DAE) represents an extension of the standard autoencoder. While the AE is designed to reconstruct input data, the DAE, in contrast, is tasked with reconstructing the original data from noisy input. In essence, the encoder of the DAE processes noisy data denoted as x'', and the decoder reconstructs the corresponding noise-free data, which is presented as x. It is important to note that apart from this distinction in their reconstruction objectives, both the AE and DAE share an identical model construction. The input data of DAE adds some extra noise, which could be random noise(Vincent et al., ; Vincent et al., 2008).

Figure 1 illustrates our proposed DAE, delineating two distinct components: one for training the DAE and the other one for imputation of missing values. Initially, the dataset was partitioned into two sub-datasets: one is samples with missing values, and the other contains samples with non-missing values.

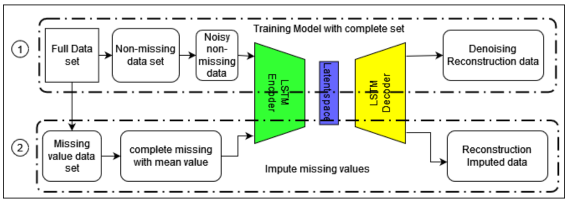

Figure 1: Denoising LSTM autoencoder imputation architecture.

The non-missing dataset adds some noise as mean values in the first block. The mean value replaces the non-missing dataset's exact location where data are missing in the missing value dataset. So, the encoder, decoder and the loss function of the DAE model are expressed as follows: $z = f(w \times x'' + b)$, $x = g(w' \times z + b')$ and $L(x, g(f(x'')))$.

The second block of Figure 1 depicts the imputation procedure. In this process, complete the missing value with the mean value as noise in the missing dataset. After training the model with the noisy complete dataset. Now, test the missing values for imputation. The reconstruction data present imputation data. We construct the model as simply as possible because of the small dataset and reduce the overfitting.

4.3 LSTM Autoencoder

The denoising LSTM autoencoder is an extension of DAE (Coto-Jimenez et al., 2018). Figure 1 depicts an alternative DAE model, specifically built as a Denoising Long Short-Term Memory Autoencoder (LSTM DAE). It's important to note that the primary distinction between the first and second models lies in the design of the encoder and decoder. The initial DAE model employed conventional neurons for its autoencoder, while the LSTM DAE model was constructed using LSTM components. Notwithstanding this disparity, all other configurations and parameters of the models remain consistent. Similarly, the second block, corresponding to missing value imputations, follows the same procedure. Another consideration pertains to data preparation for LSTM; details on this aspect are provided in the data preparation section.

5 DATA DESCRIPTION AND PREPARATION

5.1 Dataset

In this research study, we harnessed emergency medical data, which is not publicly accessible, and aimed to address the challenge of missing value imputation. The data originated from emergency medical cases involving numerous patients, thus constituting a multifaceted dataset encompassing various attributes and parameters, including categorical, numerical, and text-based data. This dataset exhibited substantial heterogeneity. Our focus primarily centred on numerical data, such as blood pressure, oxygen levels, and various measurements, which are inherently specific to individual diseases. The dataset encompasses approximately 200 distinct diseases with varying prevalence. While some diseases are commonly encountered, others are quite rare, resulting in a pronounced class imbalance within the dataset.

The dataset, comprising about 16,000 samples, has an incidence of approximately 50% missing values. Of these samples, 9,834 are complete, and the rest exhibit missing data. The missing values appear randomly, with little discernible interrelationship among variables, suggesting a lack of inherent associations. These omissions result from the data generation process, where telenotary doctors input information during emergency cases via telephone and video calls without mandatory field checks or reminder functions. A synthetic dataset mirroring the actual distribution is available in the appendix.

5.2 Data Preparation

At first, we separated the dataset into two parts: missing and non-missing datasets. To train a denoising autoencoder, we need one noisy train dataset, which will be reconstructed into original data. Traditionally, extra noise in the training dataset and train the model. We add the noise in a slightly different way, not using random noise on the training dataset.

Figure 2 demonstrates the process of adding noise. In this figure, there are two blocks presents. The first block shows a simple block diagram of how the process is realized. The second block presents the same as the first block but in the form of a table representation of missing imputation. At first, we have to identify the location of the missing values on missing samples. Then, we place the exact location on the complete dataset and replace the mean of the non-missing dataset. Now, this non-missing dataset gets corrupted or noisy(Ma et al., 2020).

We employed various missing data percentages for imputation within the dataset. These missing data percentages encompassed values of 20%, 30%, 40%, and 50%.

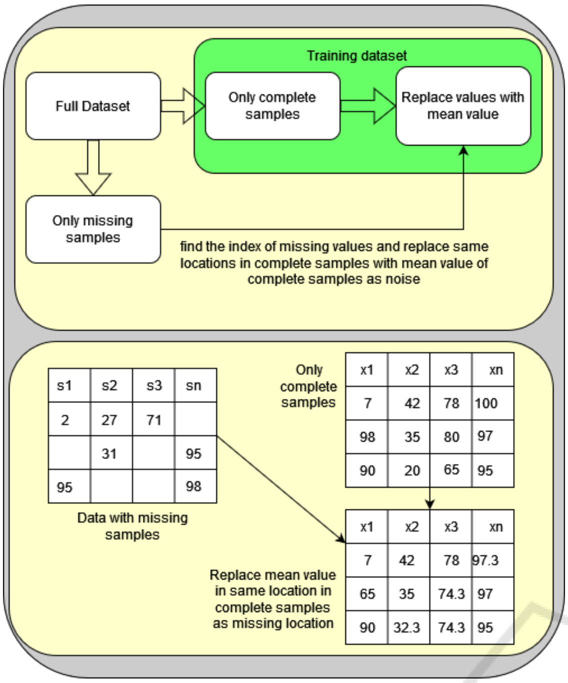

Figure 2: Data preparation for denoising autoencoder.

6 EXPERIMENTS

This experiment section describes two model configurations: the activation function, the evaluation matrix and the results.

6.1 Model Configuration

From the complete dataset with no missing values, 12% was assigned for creating the ground truth, and the rest constituted the training dataset. Within the training dataset, 70% was used for training, and the remaining 30% served for validation. To optimize results, we adopted a trial-and-error approach, leading to a variable training epoch. Given the highly class imbalance and presence of outliers, we implemented early stopping to address potential overfitting.

Table 1: Model hyperparameter and hyperparameter range.

Parameter	value range
Learning rate	$0.0001 \leq n \leq 0.01$
Batch size	$16 \leq n \leq 64$
Hidden layers	$0 \leq n \leq 2$
Hidden layers units	$16 \leq n \leq 64$
Activation function	Relu and Softmax

Table 1 outlines the hyperparameters and their corresponding ranges utilized in both models, carefully selected based on superior performance compared to alternative configurations. Key hyperparameters influencing model construction include hidden layers, units within hidden layers, activation functions, learning rate, and batch size. Learning rates spanned from 0.0001 to 0.01, with batch sizes ranging from 16 to 64. An extended exploration considered learning rates from 0.00001 to 0.1 and batch sizes from 4 to 512, with larger batch sizes incorporating batch size normalization layers. Hidden layer parameters were optimized through exploration, varying from 0 to 2 layers and an additional experiment ranging from 0 to 5 layers. Hidden layer units ranged from 16 to 64, with an extensive analysis considering units from 4 to 512. The Relu activation function was predominantly used in hidden layers, while softmax was exclusively employed in the dense output layer. Various activation functions were evaluated for their efficacy in hidden layer presentations. The mean squared error (MSE) loss function served as the training metric, assessing loss across both training and validation datasets. This paper introduces two models with 24 distinct hyperparameter combinations for individual missing percentages, each trained separately. The identified hyperparameter ranges are crucial for model training and missing imputation.

6.2 Evaluation Matrix

In the test experiment, several evaluation matrices were performed to test the ground truth and reconstruction of ground truth. Other measures have also been analysed, but the RSME provides the most reliable scores. This paper includes only root mean square error (RMSE). The RMSE is defined as

$$\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

In the equation above, n refers to the number of observations. $\hat{y}_i$ and y_i are presents as predicted and original data respectively (Tyagi et al., 2022).

6.3 Results

In this section, we analyze the experiment outcomes, emphasizing the optimal performance achieved with diverse hyperparameters. Table 2 exclusively showcases the top result among 24 unique model configurations. For each missing percentage category (e.g., 20%, 30%, etc.), the best-performing model is highlighted from the 24 configurations.

Overall, eight results were displayed from 96 models. So, we present only two figures for ground truth and predicted ground truth. We choose for 30% missing imputation results and comparison.

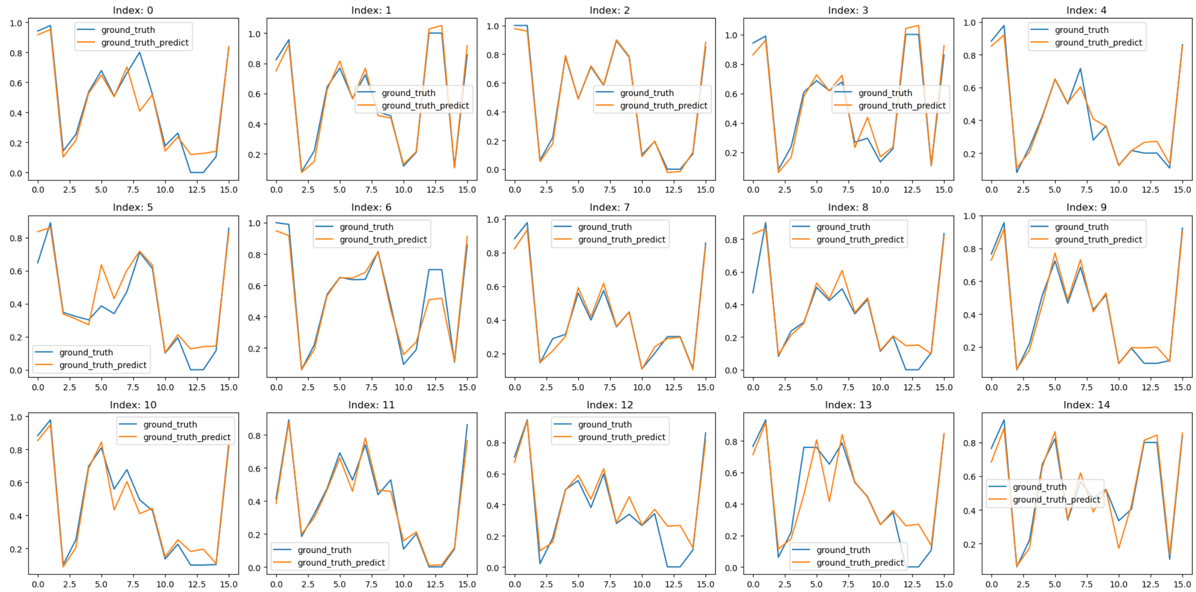

Figure 3: Comparision of ground truth and predicted ground truth for LSTM DAE.

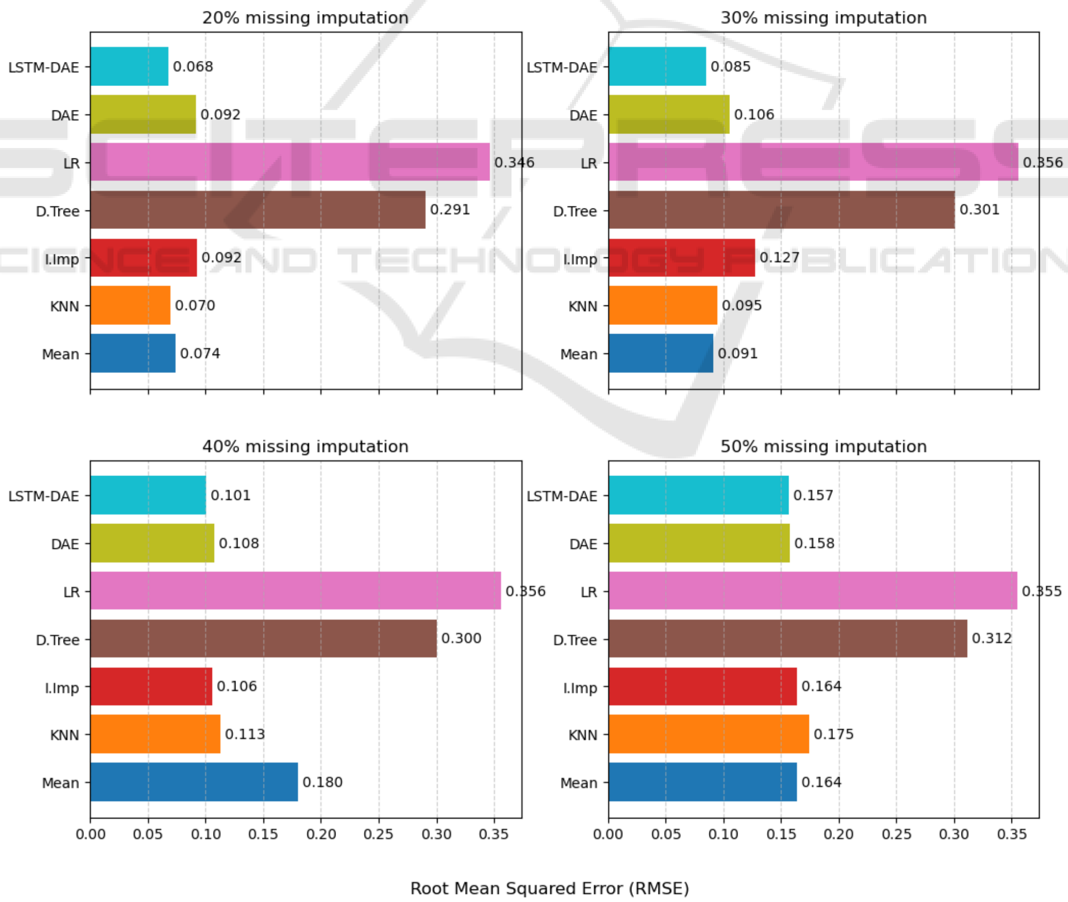

Figure 4: Comparision of seven different missing imputation methods.

Table 2: LSTM DAE model compared with several machine learning and statistical methods for missing imputation. Here is the RMSE value, which is compared to the ground truth generated.

model \missing	20%	30%	40%	50%
Mean	0.07369	0.09149	0.18007	0.16381
KNN	0.06960	0.09472	0.11314	0.17491
I.Imp	0.09231	0.12739	0.10607	0.14498
D.Tree	0.29109	0.30084	0.30034	0.31181
LR	0.34629	0.35624	0.35604	0.35461
DAE	0.09180	0.10596	0.10763	0.15755
LSTM DAE	0.06792	0.08523	0.10083	0.15702

Figure 3 illustrates the comparison between the actual ground truth and the predicted values based on the ground truth. The figure comprises 15 subplots, with each subplot representing distinct features. In Figure 3, the original ground truth is depicted by blue lines, while the orange lines indicate the predicted values.

7 DISCUSSION

The results section presents the outcomes of the missing values imputation across various missing percentage scenarios, ranging from 20% to 50% missing. Notably, the LSTM DAE consistently outperforms the standard DAE in every missing data case. The proposed models perfectly capture the class-imbalanced data.

The optimal DAE model for 20% missing imputation features a single hidden layer with 64 units for both the encoder and decoder, outperforming other configurations with a learning rate of 0.0001 and a batch size of 64. Similarly, for 30%, 40%, and 50% missing imputation, the DAE model performs best with the same configuration, adjusting the learning rate slightly to 0.001. A lower learning rate enhances model accuracy as the missing imputation increases.

For the LSTM DAE, the rise in missing values corresponds to a reduction in the number of hidden layer units. Conversely, an increase in learning rate and batch size improves performance as missing values decrease. Specifically, configurations with 64 units for the hidden layer, a learning rate of 0.0001, and a batch size of 16 exhibit superior performance for 20%, 30%, and 40% missing values compared to all other setups. Conversely, the model for 50% missing values imputation outperforms when configured with 32 hidden layer units, a learning rate of 0.001, and a batch size of 32.

Both models demonstrate their peak performance when the proportion of missing values remains below 30%; beyond this threshold, their efficacy declines. Importantly, LSTMs, typically applied to time series data, prove effective for numerical data even without a temporal dimension, surpassing the performance of traditional DAE models in such scenarios.

Moreover, we conducted a comparative analysis of several machine-learning models, which are shown in Figure 4. Both Mean imputation and KNN demonstrate a performance proximity to LSTM-DAE when tasked with imputing 20% and 50% missing values. Conversely, DAE yields results comparable to the Iterative Imputer (I.Imp) technique. For the imputation of 30% missing values, Mean and KNN techniques exhibit similar performance to LSTM-DAE, with KNN demonstrating a decrease in accuracy as missing values escalate. Throughout all scenarios, Decision Tree (D.Tree) and Linear Regression (LR) consistently exhibit lower performance relative to the other techniques. Furthermore, LSTM-DAE consistently outperforms all other models in all instances of missing data.

In this experiment, we identified limitations in both data preprocessing and the employed models. In the data preprocessing stage, mean values were introduced as noise at the locations of missing values within the complete dataset. However, this approach may have limitations when the missing dataset is larger than the non-missing dataset. It proves effective only when the missing value dataset is consistently smaller than the non-missing dataset.

8 CONCLUSIONS

Imputing missing data in emergency medical datasets presents a significant challenge, especially when dealing with class imbalance and outliers, where missing values occur randomly. To tackle this, we utilized two models: the denoising autoencoder and the denoising LSTM autoencoder. Optimal hyperparameter tuning is essential to effectively address class imbalance during imputation. We compare our proposed models with various machine learning models, and our results demonstrate that the denoising LSTM autoencoder surpasses all other models in this context.

Future investigations will explore additional deep-learning models, such as variational autoencoder, generative adversarial network, and transformer, for imputing missing values in both categorical and continuous numerical data. The resulting imputed datasets will be utilized in multi-label classification tasks to assess the effectiveness of various imputation techniques, with classification performance serving as the evaluation metric.

ACKNOWLEDGEMENTS

This work was funded by the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung) as part of the project "KI-unterstützter Telenotarzt (KIT2)" under grant number 13N16402. We would like to thank them for their support.

REFERENCES

Aachen, U. (2023). https://www.telenotarzt.de/. *Telenotarzt*.

Abiri, N., Linse, B., Edén, P., and Ohlsson, M. (2019). Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems. *Neurocomputing*, 365:137–146.

Adhikari, D., Jiang, W., Zhan, J., He, Z., Rawat, D. B., Aickelin, U., and Khorshidi, H. A. (2022). A Comprehensive Survey on Imputation of Missing Data in Internet of Things. *ACM Computing Surveys*, 55(7):133:1–133:38.

Chang, Y.-W., Natali, L., Jamialahmadi, O., Romeo, S., Pereira, J. B., Volpe, G., and Initiative, f. t. A. D. N. (2022). Neural network training with highly incomplete medical datasets. *Machine Learning: Science and Technology*, 3(3):035001. Publisher: IOP Publishing.

Chen, S. and Guo, W. (2023). Auto-encoders in deep learning—a review with new perspectives. *Mathematics*, 11(8).

Coto-Jimenez, M., Goddard-Close, J., Di Persia, L., and Leonardo Rufiner, H. (2018). Hybrid Speech Enhancement with Wiener filters and Deep LSTM Denoising Autoencoders. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, pages 1–8.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., and Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, 8(1):140.

Hameed, W. M. and Ali, N. A. (2023). Missing value imputation techniques: A survey. *UHD J. Sci. Technol.*, 7(1):72–81.

Ho, N.-H., Yang, H.-J., Kim, J., Dao, D.-P., Park, H.-R., and Pant, S. (2022). Predicting progression of Alzheimer's disease using forward-to-backward bi-directional network with integrative imputation. *Neural Networks*, 150:422–439.

Kabir, S. and Farrokhvar, L. (2022). Non-linear missing data imputation for healthcare data via index-aware autoencoders. *Health Care Management Science*, 25(3):484–497.

Liu, M., Li, S., Yuan, H., Ong, M. E. H., Ning, Y., Xie, F., Saffari, S. E., Shang, Y., Volovici, V., Chakraborty, B., and Liu, N. (2023a). Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques. *Artificial Intelligence in Medicine*, 142:102587.

Liu, M., Li, S., Yuan, H., Ong, M. E. H., Ning, Y., Xie, F., Saffari, S. E., Shang, Y., Volovici, V., Chakraborty, B., and Liu, N. (2023b). Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques. *Artificial Intelligence in Medicine*, 142:102587.

Lu, X., Yuan, L., Li, R., Xing, Z., Yao, N., and Yu, Y. (2023). An Improved Bi-LSTM-Based Missing Value Imputation Approach for Pregnancy Examination Data. *Algorithms*, 16(1):12. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

Ma, Q., Lee, W.-C., Fu, T.-Y., Gu, Y., and Yu, G. (2020). MIDIA: exploring denoising autoencoders for missing data imputation. *Data Mining and Knowledge Discovery*, 34(6):1859–1897.

Macias, E., Serrano, J., Vicario, J. L., and Morell, A. (2021). Novel Imputation Method Using Average Code from Autoencoders in Clinical Data. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1576–1579. ISSN: 2076-1465.

Mongia, A., Sengupta, D., and Majumdar, A. (2020). deepMc: Deep Matrix Completion for Imputation of Single-Cell RNA-seq Data. *Journal of Computational Biology*, 27(7):1011–1019. Publisher: Mary Ann Liebert, Inc., publishers.

Pan, H., Ye, Z., He, Q., Yan, C., Yuan, J., Lai, X., Su, J., and Li, R. (2022). Discrete Missing Data Imputation Using Multilayer Perceptron and Momentum Gradient Descent. *Sensors*, 22(15):5645. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

Psychogyios, K., Ilias, L., and Askounis, D. (2022). Comparison of Missing Data Imputation Methods using the Framingham Heart study dataset. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–5. arXiv:2210.03154 [cs].

Sun, Y., Li, J., Xu, Y., Zhang, T., and Wang, X. (2023). Deep learning versus conventional methods for missing data imputation: A review and comparative study. *Expert Systems with Applications*, 227:120201.

Tyagi, K., Rane, C., Harshvardhan, and Manry, M. (2022). Regression analysis. In *Artificial Intelligence and Machine Learning for EDGE Computing*, pages 53–63. Elsevier.

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning - ICML '08*, pages 1096–1103, Helsinki, Finland. ACM Press.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion.

Wang, Y., Yao, H., and Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing*, 184:232–242.