VAE-HyperNetFusion: Enhancing Hypernetwork Integration with Data Augmentation for Small Tabular Datasets

Jinming Zhang, Valerio Francesco Annese, and Chunxiao Hu*

** Correspondence: Chunxiao.Hu@glasgow.ac.uk*

SUMMARY: VAE-HyperNetFusion introduces an innovative framework leveraging Variational Autoencoders (VAE) with interpolation for data augmentation, aiming at enhancing the training of hypernetwork integration in the context of small tabular datasets. This method initially focused on a proprietary prostate cancer dataset, addressing the critical challenge of limited sample size by generating augmented data, significantly enriching the subsequent training environment of hypernetwork integration models. The efficacy of this approach was demonstrated not only by its superior performance on the specialized prostate cancer dataset but also by its validation across various external tabular datasets, establishing its generalizability and robustness across different domains. By incorporating VAE for data augmentation within the VAE-HyperNetFusion framework, the risk of overfitting —a common pitfall when handling small datasets —is substantially reduced through the introduction of a more diverse and enriched pool of training samples. This innovation extends the applicability of deep learning methods to scenarios traditionally dominated by shallow learning techniques, particularly in fields requiring precise and reliable analysis of limited data, such as medical diagnostics and personalized healthcare. Our results indicate that this framework surpasses existing state-of-the-art models in small dataset applications, highlighting the potential of VAE-HyperNetFusion as a critical advancement in using deep learning to address the challenges of small tabular datasets. This work not only provides the deep learning community with a novel approach but also sets a new benchmark for future research that aim at effective analysis and interpretation of small specialized datasets using artificial intelligence, especially in healthcare.

Key words. Variational Autoencoders, Hypernetworks, Ensemble Learning, Prostate Cancer

1. INTRODUCTION

In today's data-driven scientific era, deep learning techniques have demonstrated their powerful capabilities in image recognition, natural language understanding, video content analysis, and decision-making processes [1–4]. These success stories reinforce deep learning as the core of modern artificial intelligence research. However, its performance on traditional tabular data is relatively weak, especially when dealing with small datasets [5, 6]. The root of this challenge lies in the fact that deep learning architectures designed for handling images and language, such as Convolutional Neural Networks (CNNs) [7] and Transformers [8], are not directly applicable to the characteristics of tabular data, which lack clear spatial continuity or sequential dependency.

To address this issue, this study proposes VAE-HyperNetFusion, an innovative technique aimed at building efficient deep learning models for small-scale tabular datasets by combining Variational Autoencoders (VAEs) for data augmentation and hypernetwork technology. Our method first utilizes VAEs to generate high-quality augmented instances, thereby overcoming the inherent limitations of small datasets. Subsequently, through hypernetwork technology, we further develop a network ensemble strategy that dynamically generates target network parameters adapted to different data perspectives, thereby enhancing the model's adaptability and generalization performance on small-scale, high-dimensional datasets.

As a case study of our platform, we initially focus on a specific albeit very common scenario—a small dataset produced during a preliminary clinical study (prostate cancer in this case [9]), characterized by a limited number of samples and a low-dimensional feature space. Through application to this dataset, we demonstrate how VAE-HyperNetFusion effectively improves model performance in handling challenging small-scale tabular datasets. Additionally, across multiple datasets, VAE-HyperNetFusion achieves an average accuracy improvement of 6%, highlighting its robust performance advantage.

In comprehensive experiments, our method not only shows significantly better performance than current state-of-the-art techniques on small-scale datasets, specifically in terms of accuracy and AUC. Additionally, by conducting an in-depth analysis of VAE-HyperNetFusion, we further elucidate the intrinsic mechanisms underlying its performance and the impact of hyperparameter selection on the results.

2. Related Work

Research on small tabular datasets presents significant challenges across various academic fields, particularly in scenarios where large amounts of labeled data are difficult to obtain, such as medical image analysis, precision agriculture, and few-shot learning. Numerous methods have been proposed in recent years to enhance model performance on small datasets, primarily through data augmentation techniques and innovative model architectures.

Data augmentation is one of the key techniques for improving the generalization capability of models on small datasets. Kingma and Welling [10] proposed the VAE, a powerful generative model that can generate new data points by learning the latent representations of the data, which is especially useful in data-scarce situations. Additionally, Simard et al. introduced data interpolation techniques to further expand the training set by generating new samples between existing data points [11].

Hypernetworks provide a dynamic way to adjust the parameters of target networks, which is particularly valuable when dealing with small datasets. Ha, Dai, and Le explored the potential of hypernetworks to effectively generate weights for target networks [12]. This architecture allows the model to dynamically adjust based on the characteristics of the input data, thereby achieving good performance even with limited data.

In this field, many notable works have emerged. For instance, Zoph et al. in their study on automated model search (AutoML) demonstrated how searching for the optimal network architecture can improve performance on small datasets [13]. These works collectively open new possibilities for deep learning applications on small datasets, proving that innovative data processing and model design methods can achieve satisfactory results even in data-constrained scenarios.

In the domain of tabular data, several advanced methods have been developed to address the challenges posed by small datasets. XGBoost, proposed by Chen and Guestrin, has proven to be highly effective in handling tabular data through its scalable tree boosting system [14]. The model incorporates parallel computation, regularization, and cross-validation techniques to prevent overfitting and enhance generalization capabilities.

Another notable approach is TabPFN, a transformer-based model designed specifically for small tabular datasets. This model leverages the attention mechanism to handle relationships between data features, maintaining high accuracy and robustness even with high-dimensional data [15].

HyperTab, a hypernetwork-based deep learning method, combines the strengths of random forests and neural networks by generating an ensemble of neural networks. Each target model in the ensemble processes a specific lower-dimensional view of the data, which acts as data augmentation to virtually increase the number of training samples while preventing model overfitting [16].

AutoML has also gained significant attention for its ability to enhance model performance on small datasets. AutoML systems automate feature engineering, model selection, and hyperparameter tuning, reducing manual intervention and optimizing model configurations for the best performance even with limited data [13].

These examples highlight a broader trend in deep learning research towards developing and implementing methodologies that are not only effective in utilizing small amounts of data but also enhance the model's generalization capabilities to prevent overfitting. This direction is crucial for fields where large datasets are scarce or expensive to produce, such as specialized medical research, environmental monitoring, and materials science.

3. The VAE-HyperNetFusion Model

This study proposes a novel framework that integrates data augmentation using Variational Autoencoders (VAE), hypernetwork technology, and target network ensemble, providing an innovative solution for handling small datasets. This framework is, to our knowledge, the first to jointly employ VAE and hypernetworks, aiming to optimize data utilization, enhance prediction accuracy, and address challenges traditional methods face when dealing with limited data.

The data augmentation phase combines VAEs and data interpolation techniques to effectively generate high-quality and diverse data samples. This strategy not only expands the training sample size but also improves the model's adaptability to new data.

The hypernetwork component is primarily responsible for automatically generating the parameters (weights and biases) of the target networks. This design allows for the rapid construction and optimization of multiple sub-networks, each capturing a specific dimension of the dataset, thereby enhancing the model's representation capability.

The target network ensemble learns multiple data slices in parallel, integrating the learning outcomes of each target network. This method significantly improves the model's overall understanding of the data,

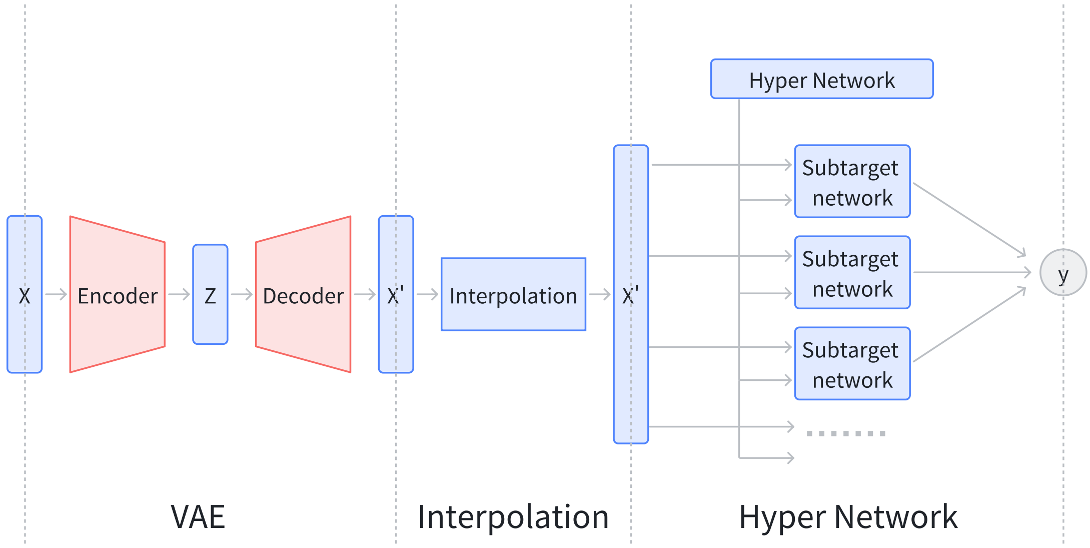

Fig. 1: This is a general VAE-HyperNetFusion architecture.

promoting an increase in prediction accuracy.

To address the challenges of limited dataset size and to ensure a fair and effective evaluation process, we implemented a Leave-P-Out cross-validation strategy with $p = 2$ during the model training phase. In this approach, we performed data augmentation exclusively on the training subsets by generating synthetic data using a Variational Autoencoder (VAE). The augmented data served as the training set for the classifier, while the left-out samples were used as the test set and remained completely unseen during both data augmentation and model training. This methodology eliminates the need to partition the scarce dataset into separate training, validation, and test sets, thereby maximizing the use of available data for training while preserving the integrity of the evaluation. The application of this framework to small datasets not only yielded superior performance but also offered a novel avenue for advancing deep learning techniques in the realm of fine-grained data analysis.

3.1. Data Augmentation

VAEs are generative models that generate new data by learning the latent representation of the input data [10]. Given original data x and latent variables z, a VAE comprises two components: the encoder $q_\phi(z|x)$ and the decoder $p_\theta(x|z)$, where ϕ and θ are the parameters of the encoder and decoder, respectively. The objective of the VAE is to maximize the Evidence Lower Bound (ELBO) of the marginal log-likelihood:

$$\begin{aligned} ELBO(\theta, \phi) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] \\ - D_{KL}(q_\phi(z|x) \| p(z)) \end{aligned} \tag{1}$$

In this equation, **E** represents the **expectation**, which measures the expected log-likelihood of the data x given the latent variable z, evaluating how well the model reconstructs the input data from the latent space. **D** denotes the **Kullback-Leibler (KL) Divergence**, which quantifies the difference between the approximate posterior $q_\phi(z|x)$ and the prior $p(z)$, serving as a regularization term to ensure the smoothness and structure of the latent space.

In this study, we employ the Variational Autoencoder (VAE) model to analyze and generate data in the medical field to explore its effectiveness in handling raw data with potential noise. We particularly focused on two key dimensions of the data to compare the differences between the original data and the data generated by the VAE. The noise points present in the original data set reflect the complexity of real medical scenarios, where individuals may have biomarkers or indicators deviating from normal values due to various diseases. This irregularity in the data, although potentially seen as processing errors at first glance, actually provides important information about the patient's health status.

Our analysis shows that the data generated by the VAE model can effectively remove these noise points while retaining the intrinsic characteristics of the original data. This finding indicates that the VAE model can not only learn and reproduce the key distribution features of the data but also reduce the impact of non-representative noise during the generation process. This is crucial for subsequent model training and data analysis as it offers an effective means to reduce the complexity of data preprocessing and improve the model's generalization ability.

By meticulously comparing the original data with the VAE-generated data, we further confirm the potential of the VAE model in the processing and analysis of medical data. This result provides significant experimental evidence for using machine learning methods to handle medical data with complex noise backgrounds and lays the foundation for future research in similar fields.

In the in-depth study of Variational Autoencoders (VAEs) and their applications in data generation and interpolation, especially when considering the generation of new data points x_{new}, we introduce an interpolation coefficient λ, mathematically defined as $\lambda \in [0, 1]$. This allows us to form a continuous path between the original data point x and the VAE-generated data point x', with the calculation formula for x_{new} as follows:

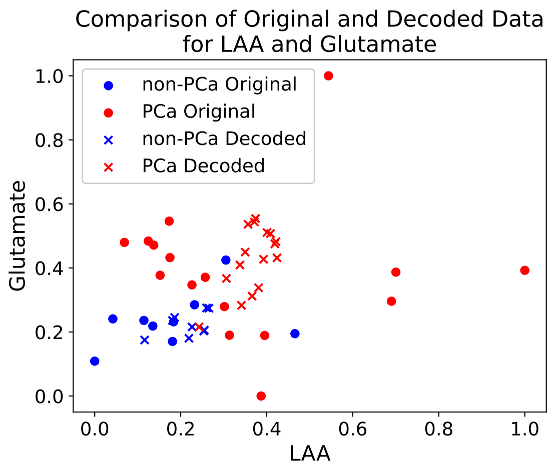

Fig. 2: Comparison of Original and Decoded Data. Note: The figure shows the data distribution and the data generated by the Variational Autoencoder (VAE). In the figure, blue circles represent data points from healthy individuals without prostate cancer, while red circles represent data points from prostate cancer patients. The cross symbols indicate data generated by the VAE model. It is observed that there are two blue data points within the red data region, which are considered noise points, representing cases of healthy individuals whose data distribution is abnormally skewed towards that of patients. After processing with the VAE, the generated data (marked by crosses) accurately capture the main characteristic distribution of the data while effectively ignoring noise points in the original data. This figure visually demonstrates the VAE model's capability in noise removal and learning the intrinsic structure of the data.

$$x_{new} = \lambda x + (1 - \lambda)x' \quad (2)$$

This formula is not only mathematically concise and elegant but also holds significant value and implications in practical applications. The following is an in-depth analysis and extended discussion of this method:

Mathematical Foundation and Implementation Details

The interpolation formula is based on the principle of linear interpolation, where λ acts as a weight to balance the contributions of the original data point and the generated data point. This balance allows us to fully obtain either the generated data point or the original data point in the extreme cases of $\lambda = 0$ or $\lambda = 1$, respectively. When λ takes a value between these two extremes, the newly generated data point x_{new} will be positioned somewhere within the data space defined by x and x', achieving a smooth transition between data points.

3.2. HyperNetFusion

Hypernetwork

The hypernetwork H_ψ is used to generate all weights and biases for the target network N_i, where ψ represents the parameters of the hypernetwork, and i indicates the index of the target network. Assuming the parameters of each target network N_i are (W_i, b_i), we have:

$$(W_i, b_i) = H_\psi(d_i) \quad (3)$$

Here, d_i is a descriptor that characterizes the features or data slices of the i-th target network.

Target Networks

Suppose there are K target networks $N_1, N_2, \ldots, N_K$, each responsible for learning a slice of the dataset. For a given input x, each target network N_k outputs a prediction y_k. The final prediction $y_{ensemble}$ is obtained by integrating the predictions of all target networks:

$$y_{ensemble} = f(y_1, y_2, \ldots, y_K) \quad (4)$$

where f is an ensemble function, which can be a simple average, weighted average, or a more complex ensemble strategy. In this paper, the basic averaging algorithm is used. Each sub-network generates a probability value for y between 0 and 1, where values greater than 0.5 are considered as class 1 (positive class), and values less than 0.5 are considered as class 0 (negative class). After applying the arithmetic average of all network outputs, the classification rule remains the same: if the average result is greater than 0.5, it is classified as a positive class; otherwise, it is classified as a negative class.

When investigating the parameter generation mechanism of the hypernetwork and target networks, an effective approach is to precisely control the data input to protect the internal structure of the hypernetwork from non-target data slices. In this study, during the training of target networks on their respective data slices, the hypernetwork should only receive data related to the current training slice, while irrelevant data should be masked (i.e., marked as false in the data input) during the training period. The core idea of this method is to ensure that the hypernetwork can generate optimal, targeted weight parameters for the target networks through fine-grained data control, while avoiding interference from unrelated data on the internal structure of the hypernetwork. This approach can improve the accuracy and robustness of the model when handling multi-task or large-scale datasets.

4. Experiments

4.1. Experimental Data

As a case study, we first focused on a small dataset obtained from a preliminary clinical trial on metabolomic detection of prostate cancer [17]. The dataset comprises the level of four plasma metabolites, namely glutamate, choline, the total concentration of L-type amino acids (LAA), and sarcosine from 10 healthy subjects (control group, non-PCa) and 16 subjects affected by prostate cancer (PCa). Data was collected using a point-of-care device performing simultaneous enzyme-based optical analysis of the sample using CMOS technology [18, 19]. The dataset collected during the prostate cancer preliminary clinical study was validated against standard measurement equipment [20]. The point-of-care platform was validated in several works and against several analytes of interest [17–26]. By using a random forest algorithm, the platform was successfully in discriminating prostate cancer with an area under the curve of 0.78 [20].

Expanding upon this single case, Table 1 provides a comprehensive comparison of multiple small datasets, each selected for its relevance in machine learning and data analysis applications. Each row represents a distinct dataset, identified by a unique number (No.), providing a clear overview of its key characteristics: name (Datasets), number of records (n), features (d), and target classes (k). For instance, the "Prostate Cancer" dataset, referenced from a specific study [17], has only 26 records but holds substantial research value. Dataset sizes vary from just 16 records in "Balloons" to a significantly larger 1484 records in "Yeast," highlighting the range of dataset scales. The diversity in features and target classes, such as the 3-feature "Haberman's Survival" dataset and the 16-feature "Zoo" dataset, reflects the variability in dimensionality and classification complexity. This variety is critical for evaluating machine learning models, as it provides a robust foundation for testing and comparing model performance across datasets of varying complexity, size, and classification challenges.

Table 1: Comparison of Characteristics Across Small Datasets. Note: Each column represents the following: **No.** - A unique identifier for each dataset; **Datasets** - The name of the dataset; **n** - The number of records in the dataset; **d** - The number of features in the dataset; **k** - The number of target classes in the dataset.

No.	Datasets	n	d	k
	Small Datasets			
0	Prostate Cancer [17]	26	4	2
1	Balloons	16	4	2
2	Lenses	24	4	3
3	Caesarian Section	80	5	2
4	Iris	150	4	2
5	Fertility	100	9	2
6	Zoo	101	16	2
7	Seeds	210	7	3
8	Haberman's Survival	306	3	2
9	Glass Identification	214	9	6
10	Yeast	1484	8	10

4.2. Experimental Models

In this study, we compare our proposed model with several well-established machine learning methods, including Random Forest (RF), XGBoost, TabPFN, HyperTab, and AutoML. Random Forest and XGBoost are popular ensemble learning algorithms, with Random Forest using multiple decision trees to improve generalization, and XGBoost employing gradient boosting to achieve superior performance on complex datasets. TabPFN and HyperTab are specifically designed for tabular data, with TabPFN combining probabilistic inference and neural networks, while HyperTab focuses on automated hyperparameter optimization and feature selection. AutoML, implemented here using TPOT, automates the machine learning pipeline to select optimal models and hyperparameters, significantly reducing model development time.

4.3. Experimental Results

From Table 2, it is evident that VAE-HNF demonstrates superior performance across various small datasets. When plotting the accuracy of different models against the parameters n, d, and k for different datasets, as illustrated in Figure 6, it is clear that the performance advantage of VAE-HNF becomes more pronounced as the dataset size decreases. Figures 7 and 8 further highlight that VAE-HNF consistently outperforms other models across different dimensions (d) and categories (k).

Table 2: Comparison of Small and Large Datasets. Numbers represent the average accuracy of five-fold cross-validation, and the numbers in parentheses are the standard deviations of the five-fold cross-validation.

Dataset	RF	XGBoost	TabPFN	HyperTab	AutoML	VAE-HNF
			Small Datasets			
0. Prostate Cancer [17]	84.00(19.5)	76.66(15.2)	70.00(17.8)	80.00(17.1)	76.00(19.59)	**86.15**
1. Balloons	77.58(10.90)	77.50(9.21)	**81.41(12.30)**	68.08(15.57)	74.91(7.95)	67.10
2. Lenses	63.00(34)	74.00(17.4)	63.00(34)	63.00(34)	60.00(33.4)	**95.83**
3. Caesarian Section	59.23(13.49)	64.57(15.26)	60.38(12.8)	63.23(14.1)	60.47(7.61)	**73.23**
4. Iris	94.66(3.39)	95.33(2.66)	96.00(2.49)	96.00(2.49)	96.00(2.49)	**97.33**
5. Fertility	88.00(7.48)	88.00(7.48)	88.00(7.48)	88.00(7.48)	86.99(5.99)	**90.00**
6. Zoo	96.04(3.73)	95.04(5.47)	96.04(1.97)	92.04(5.12)	96.04(3.73)	**100.00**
7. Seeds	92.85(2.60)	93.33(3.80)	**95.23(1.50)**	86.19(4.09)	93.80(2.42)	93.80
8. Haberman's Survival	73.21(5.66)	72.24(4.78)	73.22(6.51)	72.22(3.06)	71.25(6.75)	**73.52**
9. Glass Identification	78.93(9.96)	77.54(11.32)	71.96(6.41)	46.73(3.31)	73.84(10.51)	**84.11**
10. Yeast	61.72(3.02)	62.12(2.87)	59.49(2.85)	42.05(4.00)	61.65(3.42)	**62.73**

To calculate accuracy, we employed a Leave-P-Out cross-validation strategy, setting $p = 2$. This approach ensures that two samples are left out in each iteration for testing, while the remaining samples are used to train the model. This method provides a robust estimate of accuracy, particularly for small datasets, by maximizing the use of limited data while avoiding overfitting.

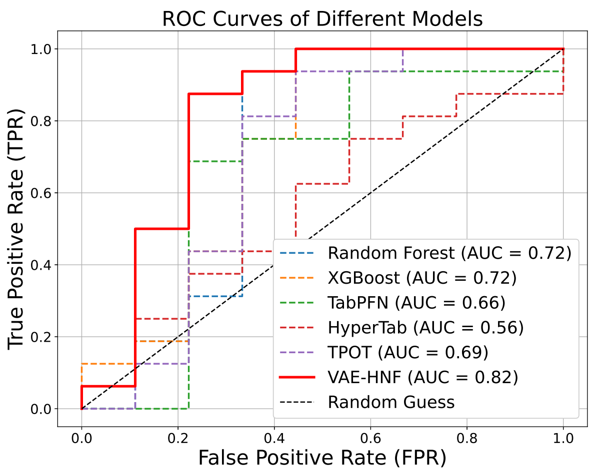

Fig. 3: ROC curves and AUC values for various models. The VAE-HNF model, highlighted in red, achieves an AUC of 0.82.

In Figure 3, the Receiver Operating Characteristic (ROC) curves visually compare the performance of the models evaluated. Notably, the VAE-HNF model exhibits the highest Area Under the Curve (AUC) at 0.82, underscoring its superior ability to distinguish between classes, particularly in small datasets. Other models, such as Random Forest and XGBoost, perform reasonably well with AUC scores of 0.72 but do not match the effectiveness of VAE-HNF. Conversely, the HyperTab model shows a lower AUC of 0.56, indicating a weaker classification capacity. The ROC curves illustrate the clear advantage of VAE-HNF in classification accuracy and reliability, emphasizing its effectiveness in scenarios with limited data.

As illustrated in Figure 4, the precision-recall curves offer a detailed view of each model' s performance, particularly under imbalanced data conditions. The VAE-HNF model demonstrates outstanding stability with near-optimal precision and recall across thresholds, outperforming all other models. Random Forest and XGBoost exhibit balanced yet moderate performance, with precision decreasing gradually as recall increases. Models like HyperTab and TPOT show more erratic patterns, with HyperTab experiencing sharp declines in precision at several points, indicating inconsistency in classification accuracy. These results underscore the robustness of the VAE-HNF model, confirming its advantage in tasks where high classification quality is critical despite data imbalance.

Figure 5 provides an in-depth comparison of various machine learning models' performance under increasing noise levels, measured by standard deviation. The VAE-HNF model shows remarkable stability, retaining high accuracy even as noise intensifies, which underscores its robustness to data imperfections. Random Forest and XGBoost, however, exhibit a sharp decline in accuracy as noise levels rise, indicating their greater vulnerability to noisy data. HyperTab and TabPFN, while initially less stable, display an intriguing resilience at higher noise levels, with their accuracy decline slowing and eventually stabilizing. This convergence at higher noise levels highlights an unexpected robustness, suggesting these models could be useful in particularly noisy environments. TPOT shows a moderate decline initially but stabilizes under medium noise conditions. This analysis demonstrates the VAE-HNF model's superior noise tolerance while also acknowledging the relative stability of HyperTab and TabPFN under high-noise settings, making them viable choices for real-world applications involving noisy or imperfect data.

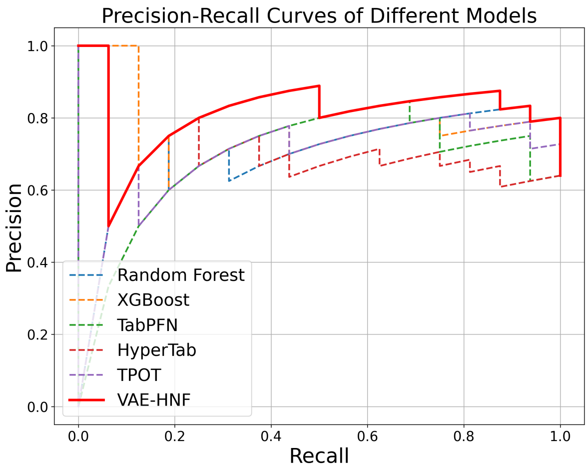

Fig. 4: Precision-Recall Curves of Different Models. The VAE-HNF model (solid red line) shows superior precision-recall performance, maintaining high precision across a range of recall values, especially in the context of imbalanced data.

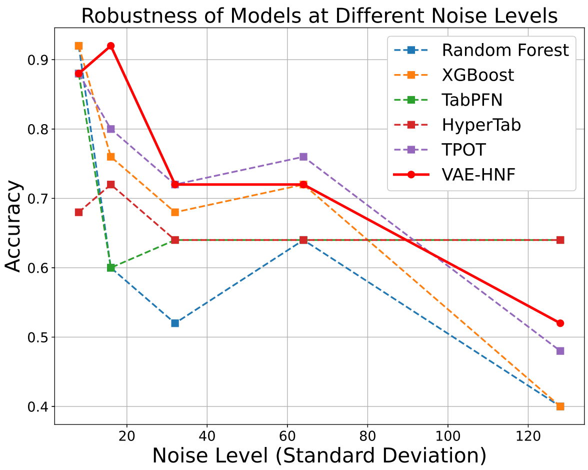

Fig. 5: Robustness of Models at Different Noise Levels. The VAE-HNF model (solid red line) maintains higher accuracy across increasing noise levels, demonstrating its robustness compared to other models.

From the Figure 6, it is evident that VAE-HNF maintains superior accuracy across various dataset sizes, particularly excelling on small datasets (with less than 100 records), where its performance is distinctly higher than other models. As the dataset

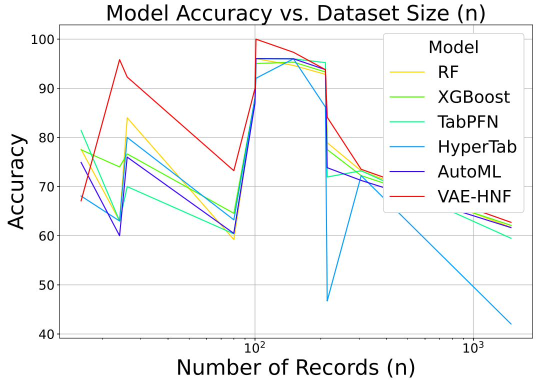

Fig. 6: Model Accuracy vs. Dataset Size. This figure shows how the accuracy of different models varies with the number of records in the dataset. The VAE-HNF model is highlighted in red for better visibility.

size increases, there is a general trend of convergence among the models, although VAE-HNF consistently stays near the top in accuracy.

Random Forest and XGBoost exhibit moderate and relatively stable accuracy as dataset sizes grow, while models like HyperTab show fluctuating performance, especially with smaller datasets, indicating sensitivity to dataset size. AutoML and TabPFN display steady improvements with increased data but still fall short compared to VAE-HNF.

This analysis underscores VAE-HNF's robustness, making it a favorable choice for scenarios where data is limited. The model's ability to retain high accuracy with smaller datasets suggests that it leverages augmented data effectively, highlighting its potential in data-constrained machine learning applications.

As shown in Figure 7, the performance of various machine learning models is evaluated against datasets with differing numbers of features (d), ranging from 4 to 16. The VAE-HNF model, indicated by a solid red line, exhibits the most stable trend, with accuracy gradually increasing as the number of features rises, showcasing its robustness and ability to handle feature-rich data effectively.

In contrast, other models, including Random Forest (RF), XGBoost, TabPFN, HyperTab, and AutoML, display more fluctuating patterns. Random Forest and XGBoost demonstrate moderate accuracy levels but experience declines at certain feature counts, indicating potential sensitivity to feature quantity. HyperTab shows a particularly volatile trend, with significant accuracy drops at specific points, suggesting it may struggle with feature complexity in these datasets. AutoML, while competitive, also shows fluctuations, though it maintains relatively high accuracy overall.

This analysis highlights the VAE-HNF model's ability to leverage increased feature data without compromising accuracy, underscoring its adaptability for datasets with varied feature dimensions. The consistent improvement in VAE-HNF's performance as features increase suggests that it may be well-suited for tasks where feature richness is critical, compared to models that show more variability and potential instability with increased feature counts.

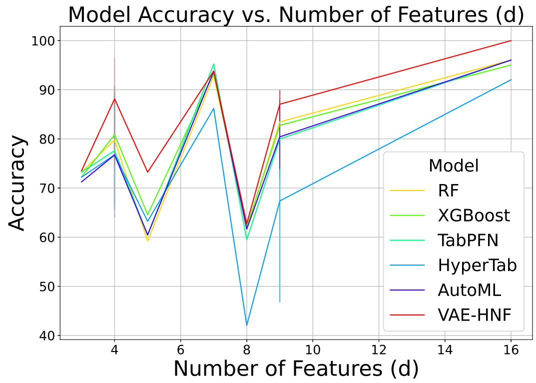

Fig. 7: Model Accuracy vs. Number of Features. This figure illustrates the relationship between model accuracy and the number of features in the dataset. The VAE-HNF model is highlighted in red.

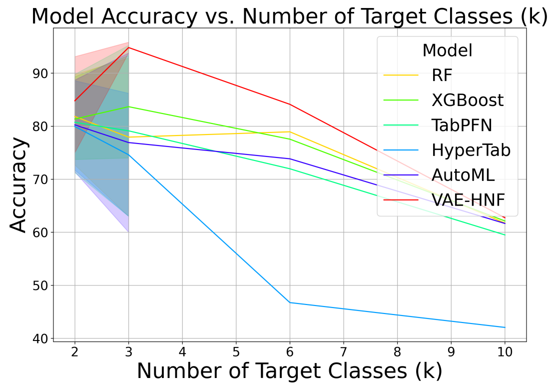

Fig. 8: Model Accuracy vs. Number of Target Classes. This figure depicts how model accuracy changes with the number of target classes. The VAE-HNF model is highlighted in red.

As illustrated in Figure 8, the accuracy trends of various models are observed as the number of target classes increases. The VAE-HNF model, represented by the solid red line, maintains a higher accuracy than other models across an increasing number of target classes, highlighting its resilience in handling complex multi-class classification tasks.

The figure shows that accuracy decreases for all models as the number of classes rises. However, the decline rate varies across models. While models like Random Forest and XGBoost display a relatively moderate decrease, HyperTab experiences a more drastic drop in accuracy, particularly beyond five classes. TabPFN and AutoML maintain moderate stability in performance but still fall behind VAE-HNF, which stands out due to its ability to retain accuracy at higher target class counts. HyperTab, by contrast, exhibits the steepest decline, underscoring its limitations in multi-class contexts.

This analysis emphasizes the suitability of VAE-HNF for applications requiring accurate classification over multiple classes, where other models tend to lose performance as complexity increases. The consistent performance advantage of VAE-HNF across varying target classes makes it a preferred model for data-rich, multi-class classification problems.

5. Conclusion and Future Work

In this study, we demonstrated the effectiveness of the VAE-HNF model across a range of small datasets, highlighting its superior performance in terms of accuracy, as shown in Table 2. The model consistently outperformed other machine learning approaches in key metrics, particularly on datasets with limited records (Figure 6). Notably, VAE-HNF maintained a high level of accuracy as dataset size decreased, underscoring its robustness in data-constrained scenarios where traditional models often underperform. Additionally, Figures 7 and 8 further illustrated VAE-HNF's stability across variations in feature dimensions and target classes, positioning it as a reliable option for tasks involving diverse and complex datasets.

Figures 3 and 4 provided insights into the model' s classification capabilities, with VAE-HNF showing the highest Area Under the Curve (AUC) in ROC and precision-recall metrics, reflecting its robust ability to distinguish between classes even in imbalanced datasets. Furthermore, VAE-HNF demonstrated exceptional resilience to data noise (Figure 5), consistently retaining accuracy at higher noise levels, which is essential for real-world applications with inherent data imperfections.

Despite these promising results, there are several directions for future research. First, applying VAE-HNF to larger and more complex datasets could provide valuable insights into its scalability and broader applicability. Moreover, exploring alternative hyperparameter configurations and incorporating different model architectures may yield deeper insights into the factors driving VAE-HNF's performance. Additionally, further experimentation with diverse data augmentation techniques could enhance its capacity to handle sample diversity, and future work might integrate VAE-HNF with advanced methods like ensemble learning or transfer learning to expand its versatility across various machine learning tasks.

In conclusion, this study highlights VAE-HNF's potential as an effective model for small dataset analysis, where conventional models often struggle. The findings from this research open multiple avenues for future investigation, including refining data augmentation strategies, enhancing model interpretability and transparency, and exploring applications in multi-task learning frameworks. These extensions hold promise for advancing VAE-HNF's utility in data-driven fields requiring robust, scalable solutions for small and noisy datasets.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[2] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[3] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[4] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

[5] I. Shavitt and E. Segal. Regularization learning networks: Deep learning for tabular datasets. *arXiv preprint arXiv:1805.06440*, 2018.

[6] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci. Deep neural networks and tabular data: A survey. *arXiv preprint arXiv:2110.01889*, 2021.

[7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30:5998–6008, 2017.

[9] M. S. Gränacher, M. Held, G. Moeller, O. Krauss, R. Zengerle, and T. Meiser. A digital twin for rapid design and virtual prototyping of microfluidic chips. *Microsystems & Nanoengineering*, 7(1):1–10, 2021.

[10] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[11] P. Y. Simard, D. Steinkraus, and J. C. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, pages 958–963, 2003.

[12] D. Ha, A. M. Dai, and Q. V. Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.

[13] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8697–8710, 2018.

[14] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[15] H. Schmidhuber and K. R. Muller. Tabpfn: A transformer that solves small tabular classification problems. *arXiv preprint arXiv:2106.03253*, 2023.

[16] Witold Wydmański, Oleksii Bulenok, and Marek Śmieja. Hypertab: Hypernetwork approach for deep learning on small tabular datasets. *arXiv preprint arXiv:2106.03253*, 2023.

[17] Valerio F. Annese, Samadhan B. Patil, Chunxiao Hu, et al. A monolithic single-chip point-of-care platform for metabolomic prostate cancer detection. *Microsyst Nanoeng*, 7:21, 2021.

[18] M. A. Al-Rawhani et al. Multimodal integrated sensor platform for rapid biomarker detection. *IEEE Transactions on Biomedical Engineering*, 67(2):614–623, 2020.

[19] Valerio F. Annese et al. Micromolar metabolite measurement in an electronically multiplexed format. *IEEE Transactions on Biomedical Engineering*, 69(9):2715–2722, 2022.

[20] Chunxiao Hu et al. Factor viii companion diagnostic for haemophilia. *Frontiers in Bioengineering and Biotechnology*, 10, 2022.

[21] Chunxiao Hu, Valerio F. Annese, Michael P. Barrett, and David R. S. Cumming. Point-of-care platform for diagnosis of venous thrombosis by simultaneous detection of thrombin generation and d-dimer in human plasma. *Analytical Chemistry*, 95(2):1115–1122, 2023.

[22] Valerio F. Annese and Chunxiao Hu. Integrating microfluidics and electronics in point-of-care diagnostics: Current and future challenges. *Micromachines*, 13:1923, 2022.

[23] Valerio F. Annese, Chunxiao Hu, Michael P. Barrett, Richard Jones, and David R. S. Cumming. A cmos-based multiomics detection platform for simultaneous quantification of proteolytically active prostate-specific antigen and glutamate in urine. *IEEE Sensors Journal*, 22(19):18870–18877, 2022.

[24] Claudio Accarino, Valerio F. Annese, Boon Chong Cheah, Mohammed A. Al-Rawhani, Yash D. Shah, James Beeley, Christos Giagkoulovitis, Srinjoy Mitra, and David R. S. Cumming. Noise characteristics with cmos sensor array scaling. *Measurement*, 152:107325, 2020.

[25] Chunxiao Hu, Valerio F. Annese, Srinivas Velugotla, Mohammed Al-Rawhani, Boon Chong Cheah, James Grant, Michael P. Barrett, and David R. S. Cumming. Disposable paper-on-cmos platform for real-time simultaneous detection of metabolites. *IEEE Transactions on Biomedical Engineering*, 67(9):2417–2426, 2020.

[26] Valerio F. Annese, Chunxiao Hu, Claudio Accarino, Christos Giagkoulovits, Samadhan B. Patil, Mohammed A. Al-Rawhani, James Beeley, et al. The multicorder: A handheld multimodal metabolomics-on-cmos sensing platform. In *2019 IEEE 8th International Workshop on Advances in Sensors and Interfaces (IWASI)*, pages 130–135, 2019.